

# Technical Report # 644: Relative Density of the Random $r$ -Factor Proximity Catch Digraph for Testing Spatial Patterns of Segregation and Association

*Elvan Ceyhan, Carey E. Priebe, & John C. Wierman*  
Department of Applied Mathematics and Statistics,  
The Johns Hopkins University, Baltimore, MD, 21218

July 29, 2004

## Abstract

Statistical pattern classification methods based on data-random graphs were introduced recently. In this approach, a random directed graph is constructed from the data using the relative positions of the data points from various classes. Different random graphs result from different definitions of the proximity region associated with each data point and different graph statistics can be employed for data reduction. The approach used in this article is based on a parameterized family of proximity maps determining an associated family of data-random digraphs. The relative arc density of the digraph is used as the summary statistic, providing an alternative to the domination number employed previously. An important advantage of the relative arc density is that, properly re-scaled, it is a  $U$ -statistic, facilitating analytic study of its asymptotic distribution using standard  $U$ -statistic central limit theory. The approach is illustrated with an application to the testing of spatial patterns of segregation and association. Knowledge of the asymptotic distribution allows evaluation of the Pitman and Hodges-Lehmann asymptotic efficacy, and selection of the proximity map parameter to optimize efficacy. Notice that the approach presented here also has the advantage of validity for data in any dimension.

## 1 Introduction

Classification and clustering have received considerable attention in the statistical literature. In recent years, a new classification approach has been developed which is based on the relative positions of the data points from various classes. Priebe et al. introduced the class cover catch digraphs (CCCD) in  $\mathbb{R}$  and gave the exact and the asymptotic distribution of the domination number of the CCCD (Priebe et al. (2001)). DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003b), Priebe et al. (2003a) applied the concept in higher dimensions and demonstrated relatively good performance of CCCD in classification. The methods employed involve data reduction (condensing) by using approximate minimum dominating sets as prototype sets (since finding the exact minimum dominating set is an NP-hard problem—in particular for CCCD). Furthermore the exact and the asymptotic distribution of the domination number of the CCCD are not analytically tractable in multiple dimensions.

Ceyhan and Priebe introduced the central similarity proximity map and  $r$ -factor proximity maps and the associated random digraphs in Ceyhan and Priebe (2003a) and Ceyhan and Priebe (2003b), respectively. In both cases, the space is partitioned by the Delaunay tessellation which is the Delaunay triangulation in  $\mathbb{R}^2$ . In each triangle, a family of data-random proximity catch digraphs is constructed based on the proximity of the points to each other. The advantages of the  $r$ -factor proximity catch digraphs are that an exact minimum dominating set can be found in polynomial time and the asymptotic distribution of the domination number is analytically tractable. The latter is then used to test segregation and association of points of different classes in Ceyhan and Priebe (2003b).

In this article, we employ a different statistic, namely the relative (arc) density, that is the proportion of all possible arcs (directed edges) which are present in the data random digraph. This test statistic has

the advantage that, properly rescaled, it is a  $U$ -statistic. Two simple classes of alternative hypotheses, for segregation and association, are defined in Section 2.5. The asymptotic distributions under both the null and the alternative hypotheses are determined in Section 3 by using standard  $U$ -statistic central limit theory. Pitman and Hodges-Lehmann asymptotic efficacy are analyzed in Sections 3.7 and 3.8, respectively. This test is related to the available tests of segregation and association in the literature, such as Pielou's test and Ripley's test. See discussion in Section 4 for more detail. Our approach is valid for data in any dimension, but for simplicity of expression and visualization, will be described for two-dimensional data.

## 2 Preliminaries

### 2.1 Proximity Maps

Let  $(\Omega, \mathcal{M})$  be a measurable space and consider a function  $N : \Omega \times \wp(\Omega) \rightarrow \wp(\Omega)$ , where  $\wp(\cdot)$  represents the power set functional. Then given  $\mathcal{Y} \subseteq \Omega$ , the *proximity map*  $N_{\mathcal{Y}}(\cdot) = N(\cdot, \mathcal{Y}) : \Omega \rightarrow \wp(\Omega)$  associates with each point  $x \in \Omega$  a *proximity region*  $N_{\mathcal{Y}}(x) \subset \Omega$ . Typically,  $N$  is chosen to satisfy  $x \in N_{\mathcal{Y}}(x)$  for all  $x \in \Omega$ . The use of the adjective *proximity* comes from thinking of the region  $N_{\mathcal{Y}}(x)$  as representing a neighborhood of points “close” to  $x$  (Jaromczyk and Toussaint (1992); Toussaint (1980)).

### 2.2 $r$ -Factor Proximity Maps

We now briefly define  $r$ -factor proximity maps (see Ceyhan and Priebe (2003b) for more details). Let  $\Omega = \mathbb{R}^2$  and let  $\mathcal{Y} = \{y_1, y_2, y_3\} \subset \mathbb{R}^2$  be three non-collinear points. Denote by  $T(\mathcal{Y})$  the triangle—including the interior—formed by the three points. For  $r \in [1, \infty]$ , define  $N_{\mathcal{Y}}^r$  to be the  $r$ -factor proximity map as follows; see also Figure 1. Using line segments from the center of mass of  $T(\mathcal{Y})$  to the midpoints of its edges, we partition  $T(\mathcal{Y})$  into “vertex regions”  $R(y_1)$ ,  $R(y_2)$ , and  $R(y_3)$ . For  $x \in T(\mathcal{Y}) \setminus \mathcal{Y}$ , let  $v(x) \in \mathcal{Y}$  be the vertex in whose region  $x$  falls, so  $x \in R(v(x))$ . If  $x$  falls on the boundary of two vertex regions, we assign  $v(x)$  arbitrarily to one of the adjacent regions. Let  $e(x)$  be the edge of  $T(\mathcal{Y})$  opposite  $v(x)$ . Let  $\ell(x)$  be the line parallel to  $e(x)$  through  $x$ . Let  $d(v(x), \ell(x))$  be the Euclidean (perpendicular) distance from  $v(x)$  to  $\ell(x)$ . For  $r \in [1, \infty)$ , let  $\ell_r(x)$  be the line parallel to  $e(x)$  such that  $d(v(x), \ell_r(x)) = rd(v(x), \ell(x))$  and  $d(\ell(x), \ell_r(x)) < d(v(x), \ell_r(x))$ . Let  $T_r(x)$  be the triangle similar to and with the same orientation as  $T(\mathcal{Y})$  having  $v(x)$  as a vertex and  $\ell_r(x)$  as the opposite edge. Then the  $r$ -factor proximity region  $N_{\mathcal{Y}}^r(x)$  is defined to be  $T_r(x) \cap T(\mathcal{Y})$ . Notice that  $r \geq 1$  implies  $x \in N_{\mathcal{Y}}^r(x)$ . Note also that  $\lim_{r \rightarrow \infty} N_{\mathcal{Y}}^r(x) = T(\mathcal{Y})$  for all  $x \in T(\mathcal{Y}) \setminus \mathcal{Y}$ , so we define  $N_{\mathcal{Y}}^\infty(x) = T(\mathcal{Y})$  for all such  $x$ . For  $x \in \mathcal{Y}$ , we define  $N_{\mathcal{Y}}^r(x) = \{x\}$  for all  $r \in [1, \infty]$ .

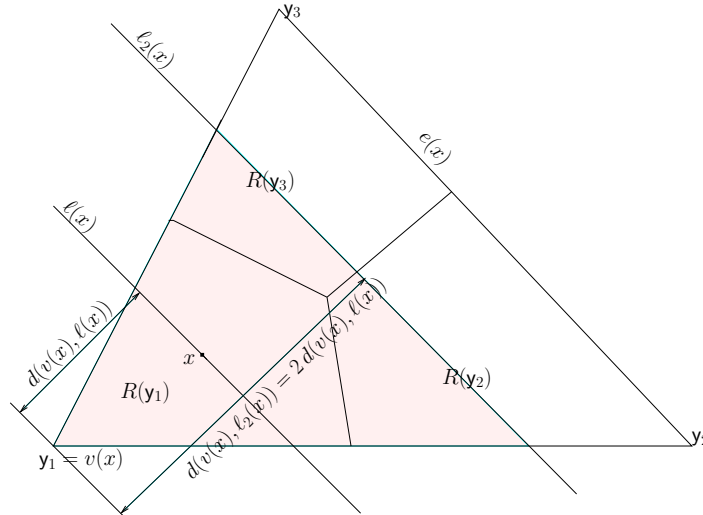


Figure 1: Construction of  $r$ -factor proximity region,  $N_{\mathcal{Y}}^2(x)$  (shaded region).

## 2.3 Data-Random Proximity Catch Digraphs

If  $\mathcal{X}_n := \{X_1, X_2, \dots, X_n\}$  is a set of  $\Omega$ -valued random variables, then the  $N_{\mathcal{Y}}(X_i)$ ,  $i = 1, \dots, n$ , are random sets. If the  $X_i$  are independent and identically distributed, then so are the random sets  $N_{\mathcal{Y}}(X_i)$ .

In the case of an  $r$ -factor proximity map, notice that if  $X_i \stackrel{iid}{\sim} F$  and  $F$  has a non-degenerate two-dimensional probability density function  $f$  with  $\text{support}(f) \subseteq T(\mathcal{Y})$ , then the special case in the construction of  $N_{\mathcal{Y}}^r - X$  falls on the boundary of two vertex regions — occurs with probability zero.

The proximities of the data points to each other are used to construct a digraph. A digraph is a directed graph; i.e. a graph with directed edges from one vertex to another based on a binary relation. Define the data-random proximity catch digraph  $D$  with vertex set  $\mathcal{V} = \{X_1, \dots, X_n\}$  and arc set  $\mathcal{A}$  by  $(X_i, X_j) \in \mathcal{A} \iff X_j \in N_{\mathcal{Y}}(X_i)$ . Since this relationship is not symmetric, a digraph is needed rather than a graph. The random digraph  $D$  depends on the (joint) distribution of the  $X_i$  and on the map  $N_{\mathcal{Y}}$ .

## 2.4 Relative Density

The *relative arc density* of a digraph  $D = (\mathcal{V}, \mathcal{A})$  of order  $|\mathcal{V}| = n$ , denoted  $\rho(D)$ , is defined as

$$\rho(D) = \frac{|\mathcal{A}|}{n(n-1)}$$

where  $|\cdot|$  denotes the set cardinality functional (Janson et al. (2000)).

Thus  $\rho(D)$  represents the ratio of the number of arcs in the digraph  $D$  to the number of arcs in the complete symmetric digraph of order  $n$ , which is  $n(n-1)$ . For brevity of notation we use *relative density* rather than relative arc density henceforth.

If  $X_1, \dots, X_n \stackrel{iid}{\sim} F$  the relative density of the associated data-random proximity catch digraph  $D$ , denoted  $\rho(\mathcal{X}_n; h, N_{\mathcal{Y}})$ , is a  $U$ -statistic,

$$\rho(\mathcal{X}_n; h, N_{\mathcal{Y}}) = \frac{1}{n(n-1)} \sum_{i < j} \sum h(X_i, X_j; N_{\mathcal{Y}}) \quad (1)$$

where

$$\begin{aligned} h(X_i, X_j; N_{\mathcal{Y}}) &= \mathbf{I}\{(X_i, X_j) \in \mathcal{A}\} + \mathbf{I}\{(X_j, X_i) \in \mathcal{A}\} \\ &= \mathbf{I}\{X_j \in N_{\mathcal{Y}}(X_i)\} + \mathbf{I}\{X_i \in N_{\mathcal{Y}}(X_j)\}. \end{aligned} \quad (2)$$

We denote  $h(X_i, X_j; N_{\mathcal{Y}})$  as  $h_{ij}$  for brevity of notation. Although the digraph is asymmetric,  $h_{ij}$  is defined as the number of arcs in  $D$  between vertices  $X_i$  and  $X_j$ , in order to produce a symmetric kernel with finite variance (Lehmann (1988)).

The random variable  $\rho_n := \rho(\mathcal{X}_n; h, N_{\mathcal{Y}})$  depends on  $n$  and  $N_{\mathcal{Y}}$  explicitly and on  $F$  implicitly. The expectation  $\mathbf{E}[\rho_n]$ , however, is independent of  $n$  and depends on only  $F$  and  $N_{\mathcal{Y}}$ :

$$0 \leq \mathbf{E}[\rho_n] = \frac{1}{2} \mathbf{E}[h_{12}] \leq 1 \text{ for all } n \geq 2. \quad (3)$$

The variance  $\mathbf{Var}[\rho_n]$  simplifies to

$$0 \leq \mathbf{Var}[\rho_n] = \frac{1}{2n(n-1)} \mathbf{Var}[h_{12}] + \frac{n-2}{n(n-1)} \mathbf{Cov}[h_{12}, h_{13}] \leq 1/4. \quad (4)$$

A central limit theorem for  $U$ -statistics (Lehmann (1988)) yields

$$\sqrt{n}(\rho_n - \mathbf{E}[\rho_n]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{Cov}[h_{12}, h_{13}]) \quad (5)$$

provided  $\mathbf{Cov}[h_{12}, h_{13}] > 0$ . The asymptotic variance of  $\rho_n$ ,  $\mathbf{Cov}[h_{12}, h_{13}]$ , depends on only  $F$  and  $N_{\mathcal{Y}}$ . Thus, we need determine only  $\mathbf{E}[h_{12}]$  and  $\mathbf{Cov}[h_{12}, h_{13}]$  in order to obtain the normal approximation

$$\rho_n \stackrel{\text{approx}}{\sim} \mathcal{N}(\mathbf{E}[\rho_n], \mathbf{Var}[\rho_n]) = \mathcal{N}\left(\frac{\mathbf{E}[h_{12}]}{2}, \frac{\mathbf{Cov}[h_{12}, h_{13}]}{n}\right) \text{ for large } n. \quad (6)$$

## 2.5 Null and Alternative Hypotheses

The phenomenon known as *segregation* involves observations from different classes having a tendency to repel each other — in our case, this means the  $X_i$  tend to be located away from all elements of  $\mathcal{Y}$ . *Association* involves observations from different classes having a tendency to attract one another, so that the  $X_i$  tend to be located near an element of  $\mathcal{Y}$ . See, for instance, Dixon (1994), Coomes et al. (1999). For statistical testing for segregation and association, the null hypothesis is generally some form of *complete spatial randomness*; thus we consider

$$H_0 : X_i \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y})).$$

If it is desired to have the sample size be a random variable, we may consider a spatial Poisson point process on  $T(\mathcal{Y})$  as our null hypothesis.

We define two simple classes of alternatives,  $H_\epsilon^S$  and  $H_\epsilon^A$  with  $\epsilon \in (0, \sqrt{3}/3)$ , for segregation and association, respectively. For  $y \in \mathcal{Y}$ , let  $e(y)$  denote the edge of  $T(\mathcal{Y})$  opposite vertex  $y$ , and for  $x \in T(\mathcal{Y})$  let  $\ell_y(x)$  denote the line parallel to  $e(y)$  through  $x$ . Then define  $T(y, \epsilon) = \{x \in T(\mathcal{Y}) : d(y, \ell_y(x)) \leq \epsilon\}$ . Let  $H_\epsilon^S$  be the model under which  $X_i \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y}) \setminus \cup_{y \in \mathcal{Y}} T(y, \epsilon))$  and  $H_\epsilon^A$  be the model under which  $X_i \stackrel{iid}{\sim} \mathcal{U}(\cup_{y \in \mathcal{Y}} T(y, \sqrt{3}/3 - \epsilon))$ . Thus the segregation model excludes the possibility of any  $X_i$  occurring near a  $y_j$ , and the association model requires that all  $X_i$  occur near a  $y_j$ . The  $\sqrt{3}/3 - \epsilon$  in the definition of the association alternative is so that  $\epsilon = 0$  yields  $H_0$  under both classes of alternatives.

**Remark:** These definitions of the alternatives are given for the standard equilateral triangle. The geometry invariance result of Theorem 1 from Section 3 still holds under the alternatives, in the following sense. If, in an arbitrary triangle, a small percentage  $\delta \cdot 100\%$  where  $\delta \in (0, 4/9)$  of the area is carved away as forbidden from each vertex using line segments parallel to the opposite edge, then under the transformation to the standard equilateral triangle this will result in the alternative  $H_{\sqrt{3\delta}/4}^S$ . This argument is for segregation with  $\delta < 1/4$ ; a similar construction is available for the other cases.

## 3 Asymptotic Normality Under the Null and Alternative Hypotheses

First we present a “geometry invariance” result which allows us to assume  $T(\mathcal{Y})$  is the standard equilateral triangle,  $T((0,0), (1,0), (1/2, \sqrt{3}/2))$ , thereby simplifying our subsequent analysis.

**Theorem 1:** Let  $\mathcal{Y} = \{y_1, y_2, y_3\} \subset \mathbb{R}^2$  be three non-collinear points. For  $i = 1, \dots, n$  let  $X_i \stackrel{iid}{\sim} F = \mathcal{U}(T(\mathcal{Y}))$ , the uniform distribution on the triangle  $T(\mathcal{Y})$ . Then for any  $r \in [1, \infty]$  the distribution of  $\rho(\mathcal{X}_n; h, N_{\mathcal{Y}}^r)$  is independent of  $\mathcal{Y}$ , hence the geometry of  $T(\mathcal{Y})$ .

**Proof:** A composition of translation, rotation, reflections, and scaling will transform any given triangle  $T_o = T(y_1, y_2, y_3)$  into the “basic” triangle  $T_b = T((0,0), (1,0), (c_1, c_2))$  with  $0 < c_1 \leq 1/2$ ,  $c_2 > 0$  and  $(1 - c_1)^2 + c_2^2 \leq 1$ , preserving uniformity. The transformation  $\phi_e : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  given by  $\phi_e(u, v) = \left(u + \frac{1-2c_1}{\sqrt{3}}v, \frac{\sqrt{3}}{2c_2}v\right)$  takes  $T_b$  to the equilateral triangle  $T_e = T((0,0), (1,0), (1/2, \sqrt{3}/2))$ . Investigation of the Jacobian shows that  $\phi_e$  also preserves uniformity. Furthermore, the composition of  $\phi_e$  with the rigid motion transformations maps the boundary of the original triangle  $T_o$  to the boundary of the equilateral triangle  $T_e$ , the median lines of  $T_o$  to the median lines of  $T_e$ , and lines parallel to the edges of  $T_o$  to lines parallel to the edges of  $T_e$ . Since the joint distribution of any collection of the  $h_{ij}$  involves only probability content of unions and intersections of regions bounded by precisely such lines, and the probability content of such regions is preserved since uniformity is preserved, the desired result follows. ■

Based on Theorem 1 and our uniform null hypothesis, we may assume that  $T(\mathcal{Y})$  is the standard equilateral triangle with  $\mathcal{Y} = \{(0,0), (1,0), (1/2, \sqrt{3}/2)\}$  henceforth.

For our  $r$ -factor proximity map and uniform null hypothesis, the asymptotic null distribution of  $\rho_n(r) = \rho(\mathcal{X}_n; h, N_{\mathcal{Y}}^r)$  can be derived as a function of  $r$ . Let  $\mu(r) := \mathbf{E}[\rho_n(r)]$  and  $\nu(r) := \mathbf{Cov}[h_{12}, h_{13}]$ . Notice that  $\mu(r) = \mathbf{E}[h_{12}]/2 = P(X_2 \in N_{\mathcal{Y}}^r(X_1))$  is the probability of an arc occurring between any pair of vertices.

### 3.1 Asymptotic Normality under the Null Hypothesis

By detailed geometric probability calculations, provided in Appendix 1, the mean and the asymptotic variance of the relative density of the  $r$ -factor proximity catch digraph can explicitly be computed. The central limit theorem for  $U$ -statistics then establishes the asymptotic normality under the uniform null hypothesis. These results are summarized in the following theorem.

**Theorem 2:** For  $r \in [1, \infty)$ ,

$$\frac{\sqrt{n}(\rho_n(r) - \mu(r))}{\sqrt{\nu(r)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (7)$$

where

$$\mu(r) = \begin{cases} \frac{37}{216}r^2 & \text{for } r \in [1, 3/2) \\ -\frac{1}{8}r^2 + 4 - 8r^{-1} + \frac{9}{2}r^{-2} & \text{for } r \in [3/2, 2) \\ 1 - \frac{3}{2}r^{-2} & \text{for } r \in [2, \infty) \end{cases} \quad (8)$$

and

$$\nu(r) = \nu_1(r)\mathbf{I}(r \in [1, 4/3)) + \nu_2(r)\mathbf{I}(r \in [4/3, 3/2)) + \nu_3(r)\mathbf{I}(r \in [3/2, 2)) + \nu_4(r)\mathbf{I}(r \in [2, \infty)) \quad (9)$$

with

$$\begin{aligned} \nu_1(r) &= \frac{3007r^{10} - 13824r^9 + 898r^8 + 77760r^7 - 117953r^6 + 48888r^5 - 24246r^4 + 60480r^3 - 38880r^2 + 3888}{58320r^4}, \\ \nu_2(r) &= \frac{5467r^{10} - 37800r^9 + 61912r^8 + 46588r^6 - 191520r^5 + 13608r^4 + 241920r^3 - 155520r^2 + 15552}{233280r^4}, \\ \nu_3(r) &= -[7r^{12} - 72r^{11} + 312r^{10} - 5332r^8 + 15072r^7 + 13704r^6 - 139264r^5 + 273600r^4 - 242176r^3 \\ &\quad + 103232r^2 - 27648r + 8640]/[960r^6], \\ \nu_4(r) &= \frac{15r^4 - 11r^2 - 48r + 25}{15r^6}. \end{aligned}$$

For  $r = \infty$ ,  $\rho_n(r)$  is degenerate.

See Appendix 1 for the proof.

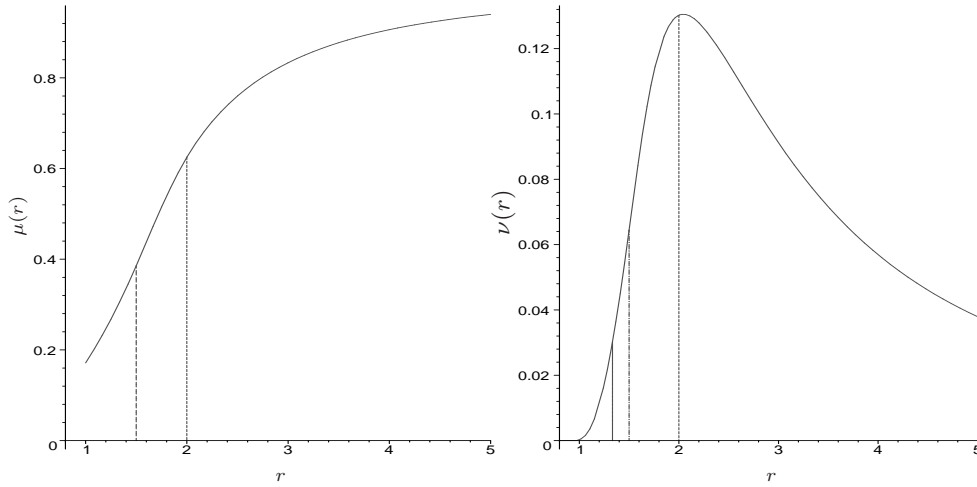


Figure 2: Asymptotic null mean  $\mu(r)$  (left) and variance  $\nu(r)$  (right) from Theorem 2. The vertical lines indicate the endpoints of the intervals in the piecewise definition of the functions. Notice that the vertical axes are differently scaled.

Consider the forms of the mean and asymptotic variance functions, which are depicted in Figure 2. Note that  $\mu(r)$  is monotonically increasing in  $r$ , since  $N_{\mathcal{Y}}^r(x)$  increases with  $r$  for all  $x \in R_{CM}(\mathbf{y}_j) \setminus \mathcal{R}_S(N_{\mathcal{Y}}^r, M_C)$ , where  $\mathcal{R}_S(N_{\mathcal{Y}}^r, M_C) := \{x \in T(\mathcal{Y}) : N_{\mathcal{Y}}^r(x) = T(\mathcal{Y})\}$ . In addition,  $\mu(r) \rightarrow 1$  as  $r \rightarrow \infty$  (at rate  $O(r^{-2})$ ), since the

digraph becomes complete asymptotically, which explains why  $\rho_n(r)$  becomes degenerate, i.e.  $\nu(r = \infty) = 0$ . Note also that  $\mu(r)$  is continuous, with the value at  $r = 1$ ,  $\mu(1) = 37/216 \approx .1713$ .

Regarding the asymptotic variance, note that  $\nu(r)$  is also continuous in  $r$  with  $\lim_{r \rightarrow \infty} \nu(r) = 0$  and  $\nu(1) = 34/58320 \approx .000583$  and observe that  $\sup_{r \geq 1} \nu(r) \approx .1305$  at  $\text{argsup}_{r \geq 1} \nu(r) \approx 2.045$ .

To illustrate the limiting distribution,  $r = 2$  yields

$$\frac{\sqrt{n}(\rho_n(2) - \mu(2))}{\sqrt{\nu(2)}} = \sqrt{\frac{192n}{25}} \left( \rho_n(2) - \frac{5}{8} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

or equivalently,

$$\rho_n(2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{5}{8}, \frac{25}{192n}\right).$$

The finite sample variance and skewness may be derived analytically in much the same way as was  $\mathbf{Cov}[h_{12}, h_{13}]$  for the asymptotic variance. In particular, the variance of  $h_{12}$  is

$$\omega(r) = \mathbf{Var}[h_{12}] = \omega_{1,1}(r) \mathbf{I}(r \in [1, 4/3)) + \omega_{1,2}(r) \mathbf{I}(r \in [4/3, 3/2)) + \omega_{1,3}(r) \mathbf{I}(r \in [3/2, 2)) + \omega_{1,4}(r) \mathbf{I}(r \in [2, \infty))$$

where

$$\begin{aligned} \omega_{1,1}(r) &= \frac{-(1369 r^8 + 4107 r^7 + 902 r^6 - 78084 r^5 + 161784 r^4 - 182736 r^3 - 23328 r^2 + 155520 r - 55296)}{11664 (r+2)(r+1)r^2}, \\ \omega_{1,2}(r) &= -\frac{1369 r^7 + 4107 r^6 + 9650 r^5 - 98496 r^4 + 132624 r^3 - 79056 r^2 - 57888 r + 72576}{11664 (r+2)(r+1)r}, \\ \omega_{1,3}(r) &= -\frac{r^{10} + 3 r^9 - 62 r^8 + 968 r^6 - 1704 r^5 - 1824 r^4 + 5424 r^3 - 1168 r^2 - 3856 r + 2208}{16 (r+2)(r+1)r^4}, \\ \omega_{1,4}(r) &= \frac{3 r^3 + 3 r^2 + 3 r - 13}{r^4(r+1)}. \end{aligned}$$

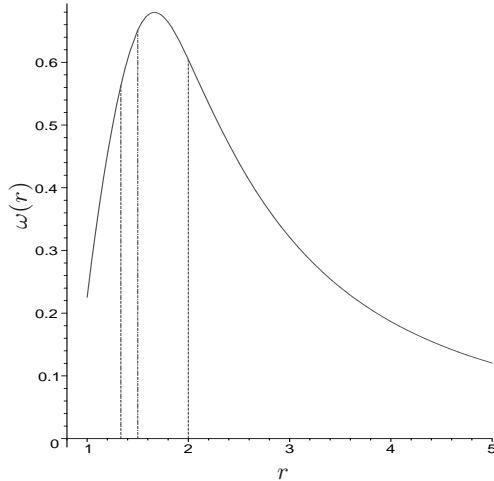


Figure 3:  $\mathbf{Var}[h_{12}] = \omega(r)$  as a function of  $r$  for  $r \in [1, 5]$ .

In Figure 3 is the graph of  $\omega(r)$  for  $r \in [1, 5]$ . Note that  $\omega(r = 1) = 2627/11664 \approx .2252$  and  $\lim_{r \rightarrow \infty} \omega(r) = 0$  (at rate  $O(r^{-2})$ ),  $\text{argsup}_{r \in [1, \infty)} \omega(r) \approx 1.66$  with  $\sup_{r \in [1, \infty)} \omega(r) \approx .6796$ .

In fact, the exact distribution of  $\rho_n(r)$  is, in principle, available by successively conditioning on the values of  $X_i$ . Alas, while the joint distribution of  $h_{12}, h_{13}$  is available, the joint distribution of  $\{h_{ij}\}_{1 \leq i < j \leq n}$ , and hence the calculation for the exact distribution of  $\rho_n(r)$ , is extraordinarily tedious and lengthy for even small values of  $n$ .

Figure 4 indicates that, for  $r = 2$ , the normal approximation is accurate even for small  $n$  (although kurtosis may be indicated for  $n = 10$ ). Figure 5 demonstrates, however, that severe skewness obtains for small values of  $n$  and extreme values of  $r$ .

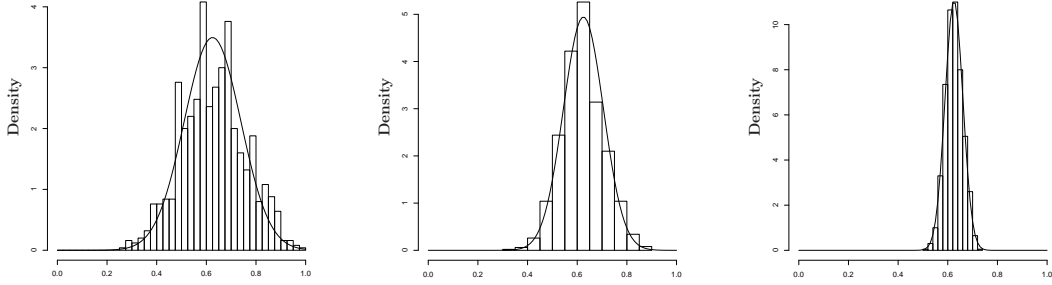


Figure 4: Depicted are the distributions of  $\rho_n(2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{5}{8}, \frac{25}{192n}\right)$  for 10, 20, 100 (left to right). Histograms are based on 1000 Monte Carlo replicates. Solid curves represent the approximating normal densities given in Theorem 2. Note that the vertical axes are differently scaled.

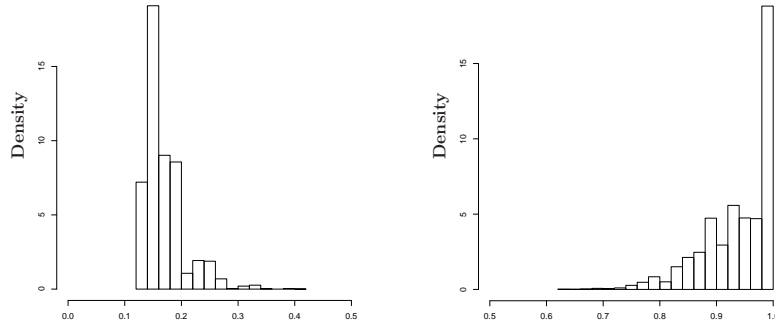


Figure 5: Depicted are the histograms for 10,000 Monte Carlo replicates of  $\rho_{10}(1)$  (left) and  $\rho_{10}(5)$  (right) indicating severe small sample skewness for extreme values of  $r$ .

Letting  $H_n(r) = \sum_{i=1}^n h(X_i, X_{n+1})$ , the exact distribution of  $\rho_n(r)$  can be written as the recurrence

$$(n+1)n\rho_{n+1}(r) \stackrel{d}{=} n(n-1)\rho_n(r) + H_n(r)$$

by noting that the conditional random variable  $H_n(r)|X_{n+1}$  is the sum of  $n$  independent and identically distributed random variables. Alas, this calculation is also tedious for large  $n$ .

### 3.2 Asymptotic Normality Under the Alternatives

Asymptotic normality of relative density of the proximity catch digraphs under the alternative hypotheses of segregation and association can be established by the same method as under the null hypothesis. Let  $\mathbf{E}_\epsilon^S[\cdot]$  ( $\mathbf{E}_\epsilon^A[\cdot]$ ) be the expectation with respect to the uniform distribution under the segregation (association) alternatives with  $\epsilon \in (0, \sqrt{3}/3)$ .

**Theorem 3** Let  $\mu_S(r, \epsilon)$  be the mean  $\mathbf{E}_\epsilon^S[h_{12}]$  and  $\nu_S(r, \epsilon)$  be the covariance,  $\mathbf{Cov}_\epsilon^S[h_{12}, h_{13}]$  for  $r \in [1, \infty]$  and  $\epsilon \in [0, \sqrt{3}/3)$  under  $H_\epsilon^S$ ,  $\sqrt{n}(\rho_n(r) - \mu_S(r, \epsilon)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_S(r, \epsilon))$  for the values of  $(r, \epsilon)$  for which  $\nu_S(r, \epsilon) > 0$ . Likewise for  $H_\epsilon^A$ .



**Sketch of Proof:** Under the alternatives, i.e.  $\epsilon > 0$ ,  $\rho_n(r)$  is a  $U$ -statistic with the same symmetric kernel  $h_{ij}$  as in the null case. Under  $H_\epsilon^S$ , the mean  $\mu_S(r, \epsilon) = \mathbf{E}_\epsilon^S[\rho_n(r)] = \mathbf{E}_\epsilon^S[h_{12}]/2$ , now a function of both  $r$  and  $\epsilon$ , is again in  $[0, 1]$ . The asymptotic variance  $\nu_S(r, \epsilon) = \mathbf{Cov}_\epsilon^S[h_{12}, h_{13}]$ , also a function of both  $r$  and  $\epsilon$ , is bounded above by  $1/4$ , as before. Thus asymptotic normality obtains provided  $\nu_S(r, \epsilon) > 0$ ; otherwise  $\rho_n(r)$  is degenerate. Likewise for  $H_\epsilon^A$ .

The explicit forms of  $\mu_S(r, \epsilon)$  and  $\mu_A(r, \epsilon)$  are given, defined piecewise, in Section ???. Sample values of  $\mu_S(r, \epsilon)$ ,  $\nu_S(r, \epsilon)$ , and  $\mu_A(r, \epsilon)$ ,  $\nu_A(r, \epsilon)$  are given in Section 3.8.1 under segregation with  $\epsilon = \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$  and in Section 3.8.2 under association with  $\epsilon = 5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21$ . Note that under  $H_\epsilon^S$ ,

$$\nu_S(r, \epsilon) > 0 \text{ for } (r, \epsilon) \in \left[1, \sqrt{3}/(2\epsilon)\right) \times \left(0, \sqrt{3}/4\right] \cup \left[1, \sqrt{3}/\epsilon - 2\right) \times \left(\sqrt{3}/4, \sqrt{3}/3\right),$$

and under  $H_\epsilon^A$ ,

$$\nu_A(r, \epsilon) > 0 \text{ for } (r, \epsilon) \in (1, \infty) \times \left(0, \sqrt{3}/3\right) \cup \{1\} \times \left(0, \sqrt{3}/12\right). \blacksquare$$

Notice that under the association alternatives any  $r \in (1, \infty)$  yields asymptotic normality for all  $\epsilon \in (0, \sqrt{3}/3)$ , while under the segregation alternatives only  $r = 1$  yields this universal asymptotic normality.

### 3.3 The Test and Analysis

The relative density of the proximity catch digraph is a test statistic for the segregation/association alternative; rejecting for extreme values of  $\rho_n(r)$  is appropriate since under segregation we expect  $\rho_n(r)$  to be large, while under association we expect  $\rho_n(r)$  to be small. Using the test statistic

$$R = \frac{\sqrt{n}(\rho_n(r) - \mu(r))}{\sqrt{\nu(r)}}, \quad (10)$$

the asymptotic critical value for the one-sided level  $\alpha$  test against segregation is given by

$$z_\alpha = \Phi^{-1}(1 - \alpha) \quad (11)$$

where  $\Phi(\cdot)$  is the standard normal distribution function. Against segregation, the test rejects for  $R > z_\alpha$  and against association, the test rejects for  $R < z_{1-\alpha}$ .

### 3.4 Consistency

**Theorem** The test against  $H_\epsilon^S$  which rejects for  $R > z_{1-\alpha}$  and the test against  $H_\epsilon^A$  which rejects for  $R < z_\alpha$  are consistent for  $r \in [1, \infty)$  and  $\epsilon \in (0, \sqrt{3}/3)$ .

**Proof:** Since the variance of the asymptotically normal test statistic, under both the null and the alternatives, converges to 0 as  $n \rightarrow \infty$  (or is degenerate), it remains to show that the mean under the null,  $\mu(r) = \mathbf{E}[\rho_n(r)]$ , is less than (greater than) the mean under the alternative,  $\mu_S(r, \epsilon) = \mathbf{E}_\epsilon^S[\rho_n(r)]$  against segregation ( $\mu_A(r, \epsilon) = \mathbf{E}_\epsilon^A[\rho_n(r)]$  against association) for  $\epsilon > 0$ . Whence it will follow that power converges to 1 as  $n \rightarrow \infty$ .

Detailed analysis of  $\mu_S(r, \epsilon)$  in Appendix 2 indicates that under segregation  $\mu_S(r, \epsilon) > \mu(r)$  for all  $\epsilon > 0$  and  $r \in [1, \infty)$ . Likewise, detailed analysis of  $\mu_A(r, \epsilon)$  in Appendix 2 indicates that under association  $\mu_A(r, \epsilon) < \mu(r)$  for all  $\epsilon > 0$  and  $r \in [1, \infty)$ . Hence the desired result follows for both alternatives.  $\blacksquare$

**Remark:** In fact, the analysis of  $\mu_S(r, \epsilon)$  and  $\mu_A(r, \epsilon)$  under the alternatives reveals more than what is required for consistency. Under segregation, the analysis indicates that  $\mu_S(r, \epsilon_1) < \mu_S(r, \epsilon_2)$  for  $\epsilon_1 < \epsilon_2$ . Likewise, under association, the analysis indicates that  $\mu_A(r, \epsilon_1) > \mu_A(r, \epsilon_2)$  for  $\epsilon_1 < \epsilon_2$ .  $\square$

### 3.5 Monte Carlo Power Analysis Under Segregation

In segregation alternatives with  $\epsilon > 0$ , we implement the above described Monte Carlo experiment for various values of  $r$  (for which  $\rho_n(r)$  is non-degenerate). Recall that  $\rho_n(r)$  is degenerate for large  $r$  at each  $\epsilon > 0$ . In



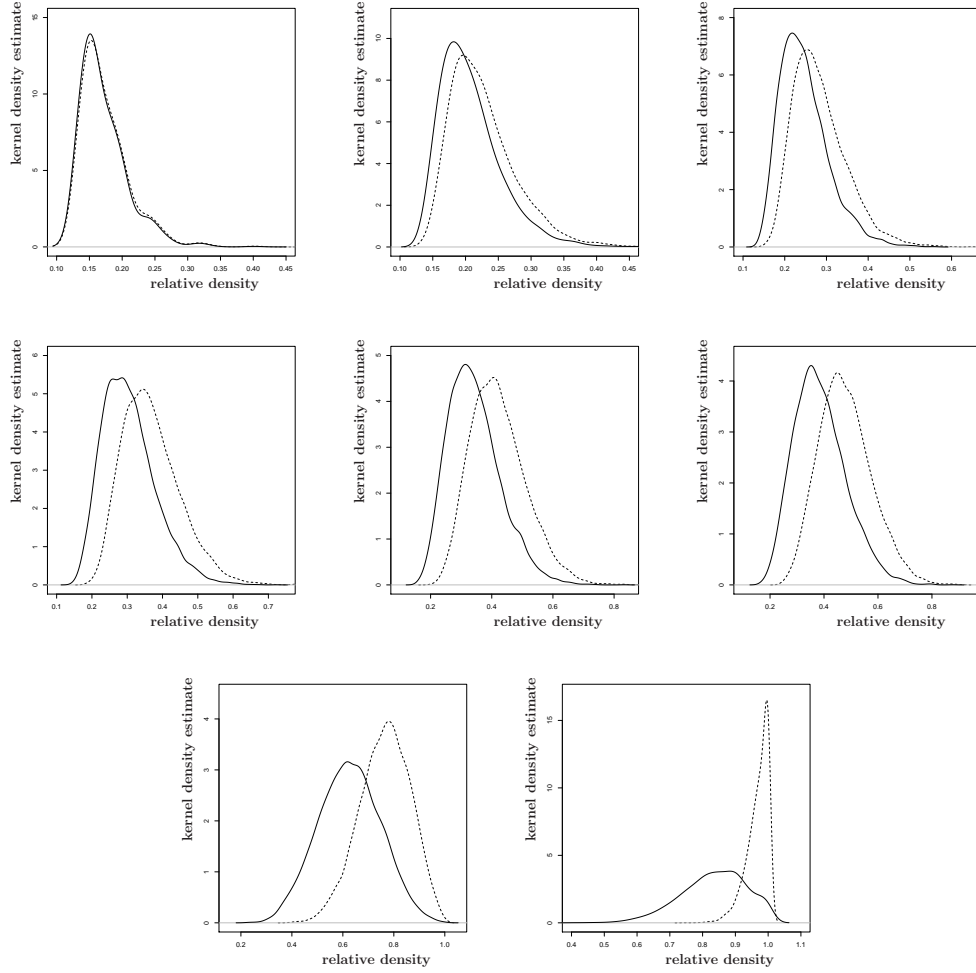


Figure 6: Kernel density estimates for the null (solid) and the segregation alternative  $H_{\sqrt{3}/8}^S$  (dashed) for  $r = 1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2, 2$ , and  $3$  (left-to-right).

particular,  $\rho_n(r, \sqrt{3}/8)$  is degenerate for  $r \geq 4$ ,  $\rho_n(r, \sqrt{3}/4)$  is degenerate for  $r \geq 2$ , and  $\rho_n(r, 2\sqrt{3}/7)$  is degenerate for  $r \geq 3/2$ .

Let  $\rho_k(n)$  be the empirical relative density for experiment  $k$  and  $\rho_{(j)}(n)$  be the  $j^{\text{th}}$  (ordered) empirical relative density for  $j = 1, \dots, N$ . Then for each  $r$  value, we estimate the empirical critical value  $\hat{C}_n^S := \rho_{(\lceil (1-\alpha)N \rceil)}(n)$  and the empirical significance level  $\hat{\alpha}_{mc}^S(n) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(\rho_j(n) > \hat{C}_n^S)$  under  $H_0$  and the empirical power  $\hat{\beta}_{mc}^S(n, \epsilon) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(\rho_j > \hat{C}_n^S)$  under  $H_\epsilon^S$  with  $\epsilon = \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$ .

For segregation with  $\epsilon = \sqrt{3}/8 \approx .2165$ , we run the Monte Carlo experiments for eight  $r$  values:  $1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2, 2$ , and  $3$ . In Figure 6 are the kernel density estimates for the null case and the segregation alternative with  $\epsilon = \sqrt{3}/8$ , for the eight  $r$  values with  $n = 10$  and  $N = 10,000$ . Observe that under both  $H_0$  and  $H_{\sqrt{3}/8}^S$ , kernel density estimates are skewed right for  $r = 1, 11/10$ , (with skewness increasing as  $r$  gets smaller) and kernel density estimates are almost symmetric for  $r = 6/5, 4/3, \sqrt{2}, 3/2, 2$ , with most symmetry occurring at  $r = 3/2$ , kernel density estimate is skewed left for  $r = 3$  (with skewness increasing as  $r$  gets larger).

The empirical critical values, empirical significance levels, and empirical power estimates under  $H_{\sqrt{3}/8}^S$  are presented in Table 1.

In Figure 7, we present a Monte Carlo investigation against the segregation alternative  $H_{\sqrt{3}/8}^S$  for  $r = 11/10$ , and  $n = 10, N = 10,000$  (left),  $n = 100, N = 1000$  (right). With  $n = 10$ , the null and alternative probability

$r$	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3
$\hat{C}_n^S$	0.24	.3	.35	.4	.5	.5	.82	0.98
$\hat{\alpha}_{mc}^S(10)$	.0324	.0403	.0484	.0442	.0446	.0492	.049	.0389
$\hat{\beta}_{mc}^S(10, \sqrt{3}/8)$	.0381	.0787	.122	.1571	.1719	.1955	.2791	.2901

Table 1: The empirical critical values, empirical significance levels, and empirical power estimates under  $H_{\sqrt{3}/8}^S$ ,  $N = 10,000$ , and  $n = 10$  at  $\alpha = .05$ .

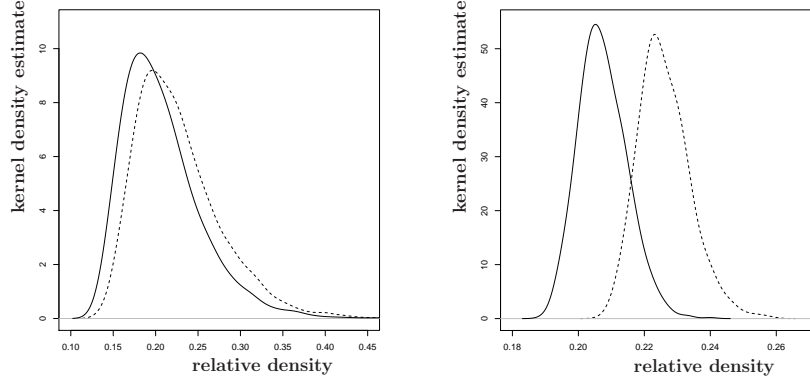


Figure 7: Two Monte Carlo experiments against the segregation alternative  $H_{\sqrt{3}/8}^S$ . Depicted are kernel density estimates for  $\rho_n(11/10)$  for  $n = 10$  (left) and  $n = 100$  (right) under the null (solid) and alternative (dashed).

density functions for  $\rho_{10}(11/10)$  are very similar, implying small power (10,000 Monte Carlo replicates yield  $\hat{\beta}_{mc}^S(10, \sqrt{3}/8) = 0.0787$ ,  $\hat{C}_n^S = 0.15$ , and  $\hat{\alpha}_{mc}^S(10) = 0.0484$ ). With  $n = 100$ , there is more separation between null and alternative probability density functions; for this case, 1000 Monte Carlo replicates yield  $\hat{\beta}_{mc}^S(100, \sqrt{3}/8) = 0.77$ ,  $\hat{C}_n^S = 0.2203$ , and  $\hat{\alpha}_{mc}^S(100) = 0.05$ . Notice also that the probability density functions are more skewed for  $n = 10$ , while approximate normality holds for  $n = 100$ .

For segregation with  $\epsilon = \sqrt{3}/4 \approx .433$ , we run the Monte Carlo experiments for six  $r$  values, 1, 11/10, 6/5, 4/3,  $\sqrt{2}$ , 3/2. In Figure 8, are the kernel density estimates for the null case and the segregation alternative with  $\epsilon = \sqrt{3}/4$ , for the six  $r$  values with  $n = 10$  and  $N = 10,000$ . Observe that under  $H_{\sqrt{3}/4}^S$ , kernel density estimate is skewed right for  $r = 1$  and kernel density estimates are almost symmetric for  $r = 11/10, 6/5, 4/3, \sqrt{2}$ , with most symmetry occurring at  $r = 4/3$ , kernel density estimate is skewed left for  $r = 3/2$ .

The empirical critical values, empirical significance levels, and empirical power estimates under  $H_{\sqrt{3}/4}^S$  are presented in Table 2.

$r$	1	11/10	6/5	4/3	$\sqrt{2}$	3/2
$\hat{C}_n^S$	.24	.3	.35	.4	.5	.5
$\hat{\alpha}_{mc}^S(10)$	.0318	.0411	.0479	.0484	.0481	.043
$\hat{\beta}_{mc}^S(10, \sqrt{3}/4)$	.1247	.9138	.998	1.0	1.0	1.0

Table 2: The empirical critical values, empirical significance levels, and empirical power estimates under  $H_{\sqrt{3}/4}^S$ ,  $N = 10,000$ , and  $n = 10$  at  $\alpha = .05$ .

For segregation with  $\epsilon = 2\sqrt{3}/7 \approx .495$ , we run the Monte Carlo experiments for six  $r$  values, 1, 21/20, 11/10, 6/5, 4/3,  $\sqrt{2}$ . In Figure 9, are the kernel density estimates for the null case and the segregation alternative with  $\epsilon = 2\sqrt{3}/7$ , for the six  $r$  values with  $n = 10$  and  $N = 10,000$ . Observe that under  $H_{2\sqrt{3}/7}^S$ , kernel density estimate is skewed right for  $r = 1$  and kernel density estimates are almost symmetric for  $r = 21/20, 11/10, 6/5$ , with most symmetry occurring at  $r = 6/5$ , kernel density estimate is skewed left for  $r = \sqrt{2}$ .

The empirical critical values, empirical significance levels, and empirical power estimates under  $H_{2\sqrt{3}/7}^S$  are

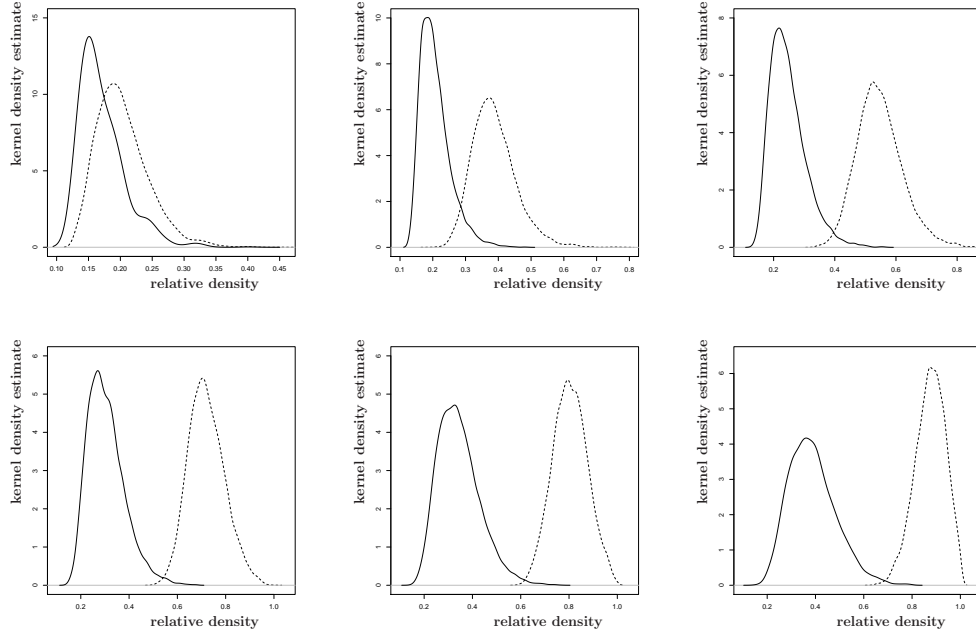


Figure 8: Kernel density estimates for the null (solid) and the segregation alternative  $H^S_{\sqrt{3}/4}$  (dashed) for  $r = 1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2$  (left-to-right).

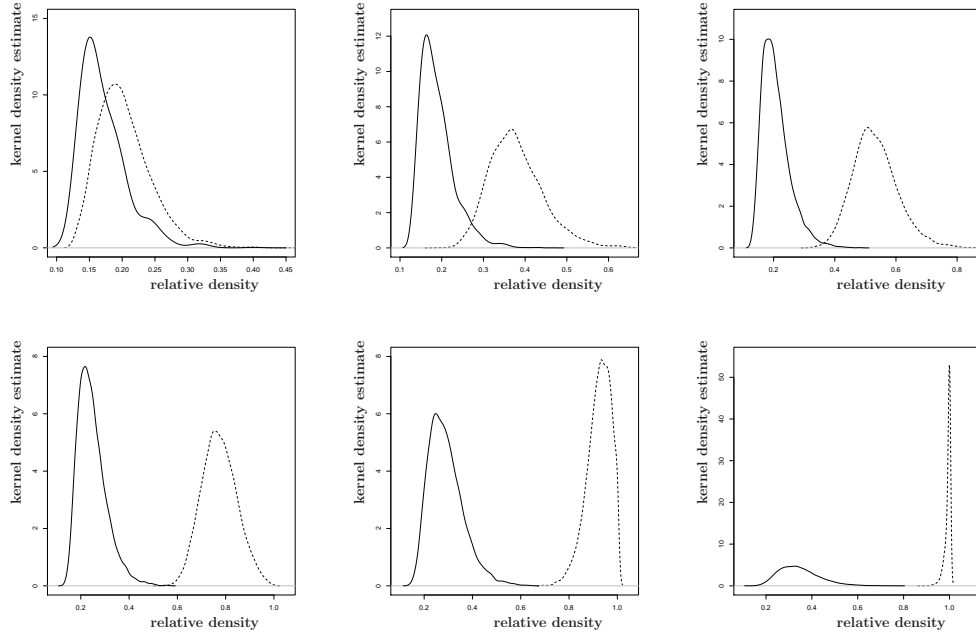


Figure 9: Kernel density estimates for the null (solid) and the segregation alternative  $H^S_{2\sqrt{3}/7}$  (dashed) for  $r = 1, 21/20, 11/10, 6/5, 4/3, \sqrt{2}$  (left-to-right).

$r$	1	21/20	11/10	6/5	4/3	$\sqrt{2}$
$\hat{C}_n^S$	.24	.28	.3	.35	.42	.5
$\hat{\alpha}_{mc}^S(10)$	.0318	.0447	.0411	.0479	.0477	.0481
$\hat{\beta}_{mc}^S(10, 2\sqrt{3}/7)$	.1247	.9728	1.0	1.0	1.0	1.0

Table 3: The empirical critical values, empirical significance levels, and empirical power estimates under  $H_{2\sqrt{3}/7}^S$ ,  $N = 10,000$ , and  $n = 10$  at  $\alpha = .05$ .

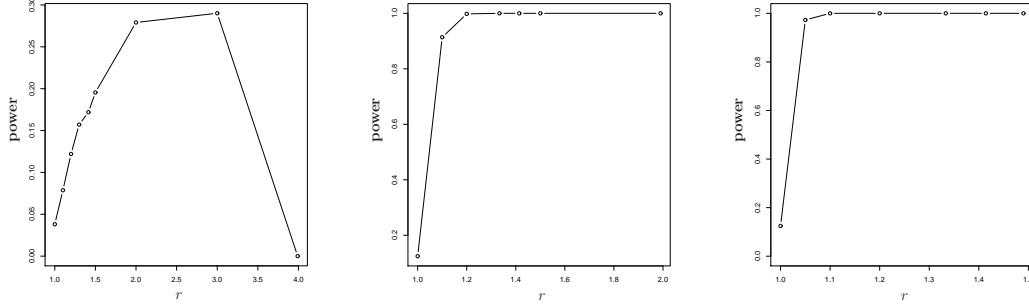


Figure 10: Monte Carlo power using the empirical critical value against segregation alternatives  $H_{\sqrt{3}/8}^S$  (left),  $H_{\sqrt{3}/4}^S$  (middle) and  $H_{2\sqrt{3}/7}^S$  (right) as a function of  $r$ , for  $n = 10$ .

presented in Table 3.

We also plot the empirical power as a function of  $r$  in Figure 10. Let  $r_S^*(\epsilon)$  be the value of  $r$  at which maximum Monte Carlo power estimate occurs, then  $r_S^*(\sqrt{3}/8) = 3$ . Furthermore, Monte Carlo power estimate increases as  $r$  gets larger and then decreases, due to the magnitude of  $r$  and  $n$ . Because for small  $n$  and large  $r$ , the critical value is approximately 1 under  $H_0$ , as we get a complete digraph with high probability.

Furthermore,  $r_S^*(\sqrt{3}/4) \in \{4/3, \sqrt{2}, 3/2\}$  and  $r_S^*(2\sqrt{3}/7) \in \{11/10, 6/5, 4/3, \sqrt{2}\}$ . Monte Carlo power estimates increase as  $r$  gets larger. The phenomenon happened above for  $\epsilon = \sqrt{3}/8$  does not occur, because  $r$  values are not large enough to yield complete digraphs under  $H_0$  with high probability.

For a given alternative and sample size, we may consider analyzing the power of the test — using the asymptotic critical value — as a function of the proximity factor  $r$ . Let  $R_j := \frac{\sqrt{n}(\rho_j(n) - \mu(r))}{\sqrt{\nu(r)}}$  be the standardized relative density for experiment  $j$  with sample size  $n$  for  $j = 1, 2, \dots, N$ . For each  $r$  value, the level  $\alpha$  asymptotic critical value is  $\mu(r) + z_{(1-\alpha)} \cdot \sqrt{\nu(r)/n}$ . We estimate the empirical power as  $\hat{\beta}_n^S(r, \epsilon) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(R_j > z_{1-\alpha})$ . In Figure 11, we present a Monte Carlo investigation of power  $\hat{\beta}_n^S(r, \epsilon)$  against  $H_{\sqrt{3}/8}^S$ ,  $H_{\sqrt{3}/4}^S$ , and  $H_{2\sqrt{3}/7}^S$  as a function of  $r$  for  $n = 10$ . The empirical significance level is  $\hat{\alpha}_{mc}^S(n) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(R_j > z_{1-\alpha} | H_0)$ . Then  $\hat{\alpha}_S(10)$ , is about .05 for  $r = 2, 3$  which have the empirical power  $\hat{\beta}_{10}^S(r, \sqrt{3}/8) \approx .35$ , and  $\hat{\beta}_{10}^S(r, \epsilon) = 1$  for  $\epsilon = \sqrt{3}/4, 2\sqrt{3}/7$ . So, for small sample sizes, moderate values of  $r$  are more appropriate for normal approximation, as they yield the desired significance level and the more severe the segregation, the higher the power estimate at each  $r$ .

The empirical significance level, and empirical power  $\hat{\beta}_n^S(r, \epsilon)$  values under  $H_\epsilon^S$  for  $\epsilon = \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$  are presented in Table 4. Note that even for  $n = 10$ , the plots of the empirical power  $\hat{\beta}_n^S(r, \epsilon)$  resemble the curves of the asymptotic power function  $\Pi_S(r)$  in Section 3.9.

### 3.6 Monte Carlo Power Analysis Under Association

In association alternatives with  $\epsilon > 0$ , we implement the Monte Carlo experiment for  $r \in \{1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2, 2, 3, 5, 10\}$ . Then for each  $r$  value, we estimate the empirical critical value  $\hat{C}_n^A := \rho_{(\lfloor \alpha N \rfloor)}$  and the

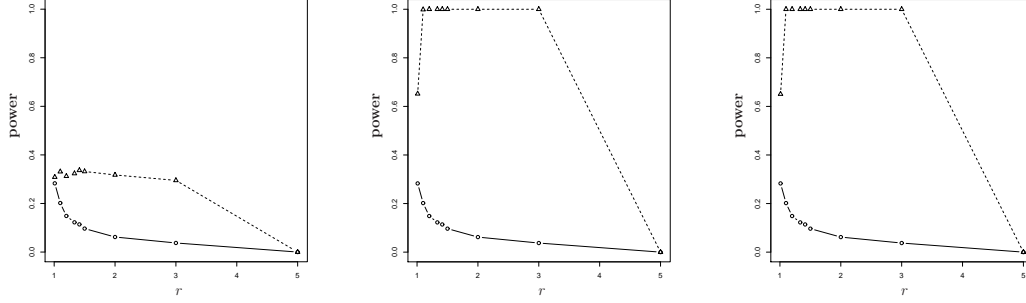


Figure 11: Monte Carlo power using the asymptotic critical value against segregation alternatives  $H_{\sqrt{3}/8}^S$  (left),  $H_{\sqrt{3}/4}^S$  (middle), and  $H_{2\sqrt{3}/7}^S$  (right) as a function of  $r$ , for  $n = 10$ . The circles represent the empirical significance levels while triangles represent the empirical power values.

$r$	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5
$\hat{\alpha}_S(n)$	.2829	.2019	.1486	.1224	.1139	.0966	.0619	.0374	.0000
$\hat{\beta}_n^S(r, \sqrt{3}/8)$	.3086	.3309	.3123	.3233	.3365	.3317	.3175	.2950	.0000
$\hat{\beta}_n^S(r, \sqrt{3}/4)$	.6519	.9985	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.0000
$\hat{\beta}_n^S(r, 2\sqrt{3}/7)$	.6508	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.0000

Table 4: The empirical significance level and empirical power values under  $H_\epsilon^S$  for  $\epsilon = \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$ ,  $N = 10,000$ , and  $n = 10$  at  $\alpha = .05$ .

empirical significance level  $\hat{\alpha}_{mc}^A(n) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(\rho_j(n) < \hat{C}_n^A)$  under  $H_0$  and the empirical power  $\hat{\beta}_{mc}^A(n, \epsilon) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(\rho_j < \hat{C}_n^A)$ . We implement the Monte Carlo simulation for three  $\epsilon$  values;  $5\sqrt{3}/24$ ,  $\sqrt{3}/12$ ,  $\sqrt{3}/21$ .

The empirical critical values, empirical significance levels, and empirical power estimates under  $H_\epsilon^A$  are presented in Table 5.

For association with  $\epsilon = 5\sqrt{3}/24 \approx .36$ , in Figure 12, are the kernel density estimates for the null case and the segregation alternative for the ten  $r$  values with  $n = 10$ ,  $N = 10,000$ . Observe that, under  $H_0$ , kernel density estimates are skewed right for  $r = 1, 11/10$ , (with skewness increasing as  $r$  gets smaller) and kernel density estimates are almost symmetric for  $r = 6/5, 4/3, \sqrt{2}, 3/2, 2$ , with most symmetry occurring at  $r = 3/2$ , kernel density estimates are skewed left for  $r = 3, 5, 10$ , (with skewness increasing as  $r$  gets larger). Under  $H_{5\sqrt{3}/24}^A$ , kernel density estimates are skewed right for  $r = 1, 11/10, 6/5, 4/3, 3/2, 2, 3$ , (with skewness increasing as  $r$  gets smaller) and kernel density estimate is almost symmetric for  $r = 5$ , kernel density estimate is skewed left for  $r = 10$ .

For association with  $\epsilon = \sqrt{3}/12 \approx .144$ , in Figure 13, are the kernel density estimates for the null case and the segregation alternative for the ten  $r$  values with  $n = 10$ ,  $N = 10,000$ . Observe that under  $H_{\sqrt{3}/12}^A$ , kernel density estimates are skewed right for  $r = 1, 11/10, 6/5, 4/3$ , (with skewness increasing as  $r$  gets smaller) and kernel density estimates are almost symmetric for  $r = \sqrt{2}, 3/2, 2$ , with most symmetry occurring at  $r = 2$ , kernel density estimates are skewed left for  $r = 3, 5, 10$ , (with skewness increasing as  $r$  gets larger).

Note also that for  $r = 11/10$  with  $n = 10$ ,  $N = 1000$ , the kernel density estimates are very similar, implying small power. With  $N = 10,000$   $\hat{\beta}_{mc}^A(10, \sqrt{3}/12) = 0.0921$ ,  $\hat{C}_n^A = 0.15$ , and  $\hat{\alpha}_{mc}^A(10) = 0.0484$ . See Figure 14. Note that for large  $n$ , there is more separation between null and alternative kernel densities, which implies higher power. With  $n = 100$ ,  $N = 1000$  and get  $\hat{C}_n^A = 0.1963$ ,  $\hat{\alpha}_{mc}^A(100) = 0.049$ , and  $\hat{\beta}_{mc}^A(100, \sqrt{3}/12) = 0.56$ .

For association with  $\epsilon = \sqrt{3}/21 \approx .0825$ , in Figure 15, are the kernel density estimates for the null case and the segregation alternative for the ten  $r$  values with  $n = 10$ ,  $N = 10,000$ . Observe that under  $H_{\sqrt{3}/21}^A$ , kernel density estimates are skewed right for  $r = 1, 11/10, 6/5, 4/3$ , (with skewness increasing as  $r$  gets smaller) and

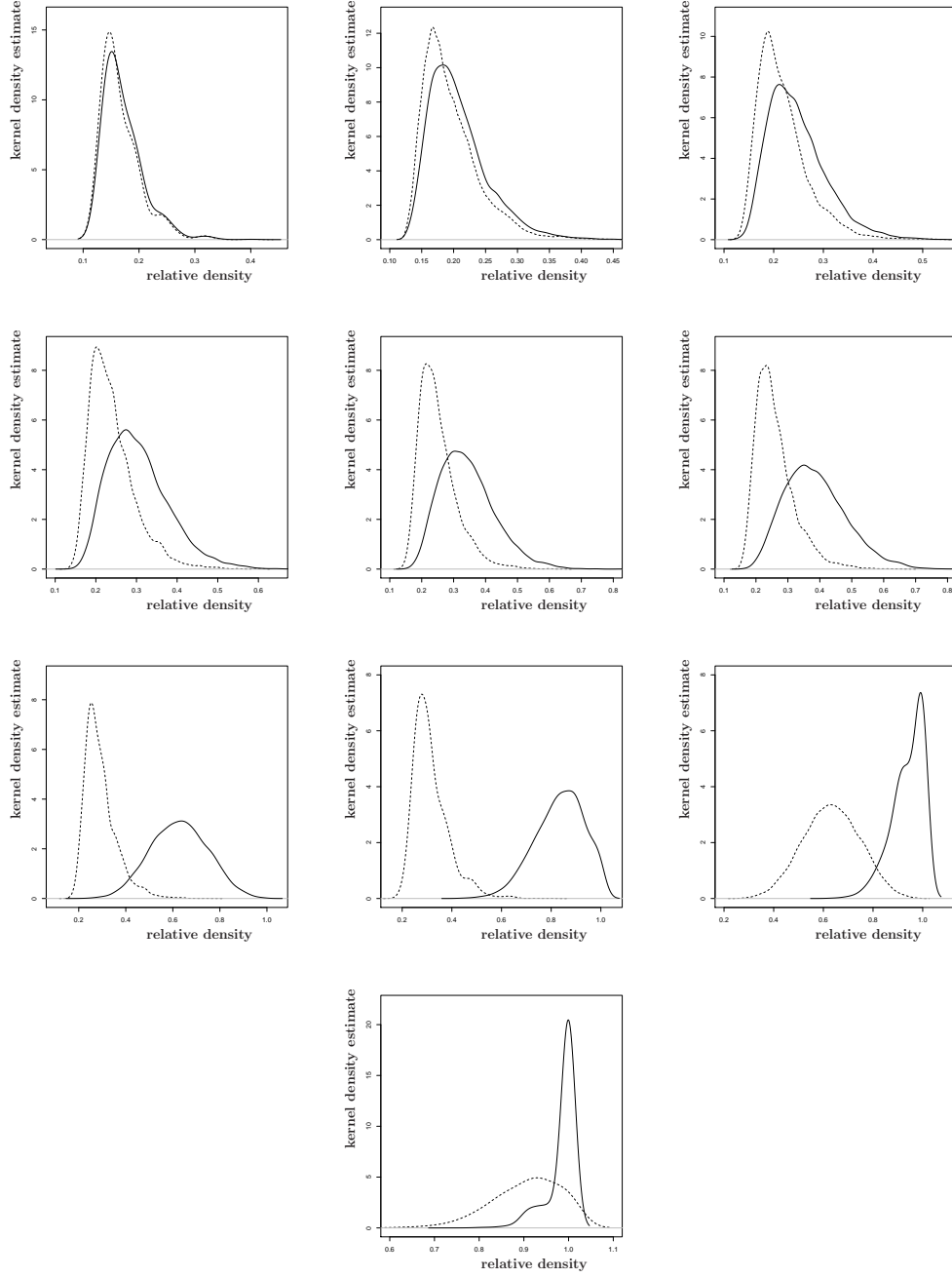


Figure 12: Kernel density estimates for the null (solid) and the association alternative  $H_{5\sqrt{3}/24}^A$  (dashed) for  $r = 1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2, 2, 3, 5, 10$  (left-to-right).

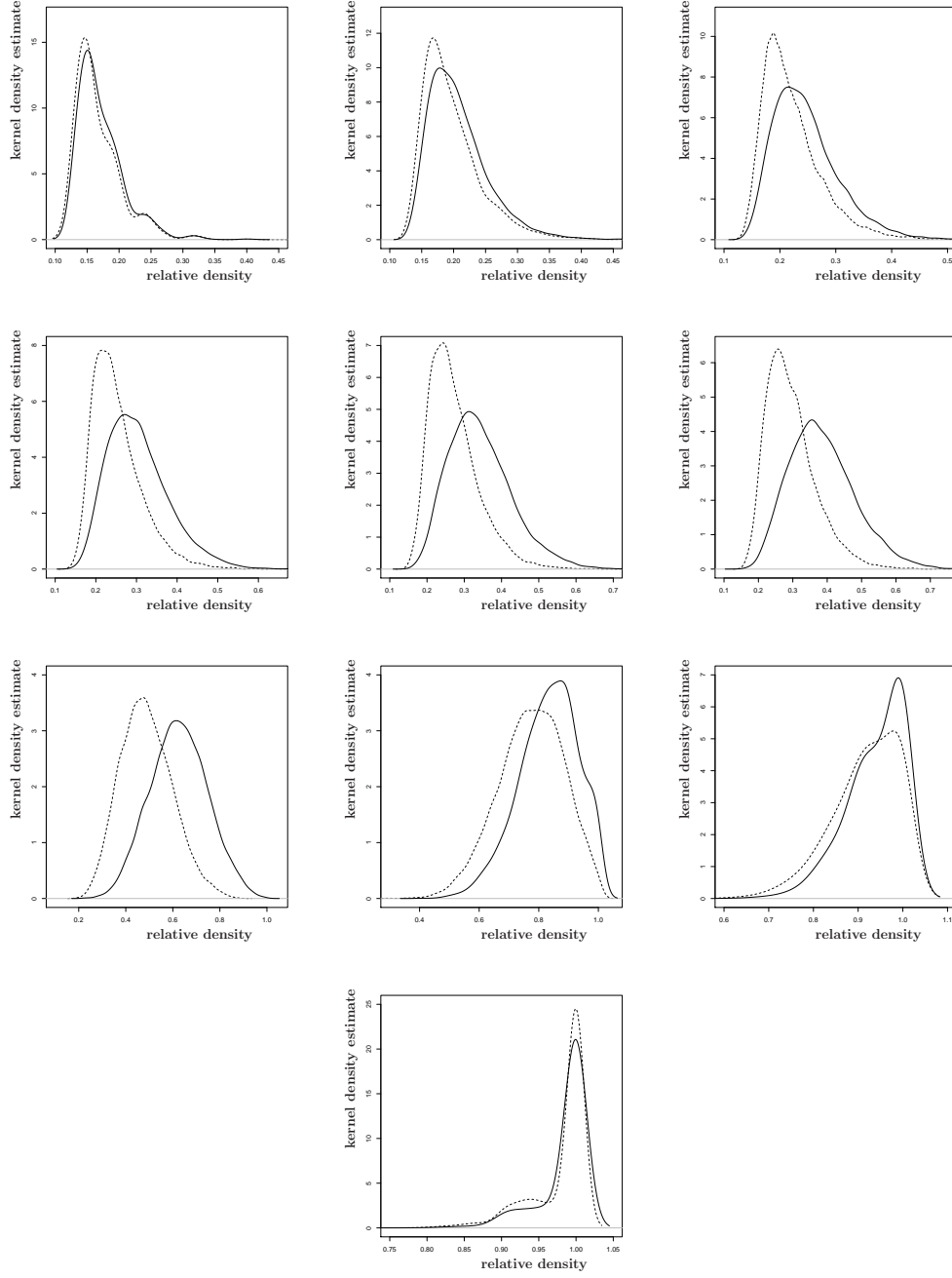


Figure 13: Kernel density estimates for the null (solid) and the association alternative  $H_{\sqrt{3}/12}^A$  (dashed) for  $r = 1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2, 2, 3, 5, 10$  (left-to-right).



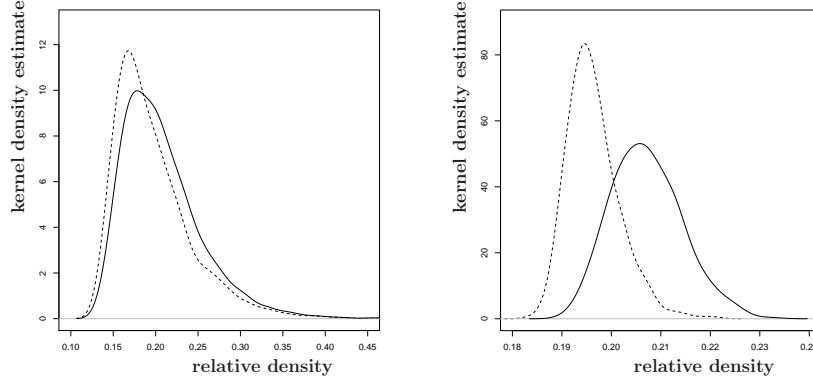


Figure 14: Two Monte Carlo experiments against the association alternative  $H_{\sqrt{3}/12}^A$ . Depicted are kernel density estimates for  $\rho_n(11/10)$  for  $n = 10$  (left) and  $n = 100$  (right) under the null (solid) and alternative (dashed).

$r$	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5	10
$\hat{C}_n^A$	.13	.14	.16	.2	.2	.24	.42	.65	.82	.91
$\hat{\alpha}_{mc}^A(10)$	0	.0112	.0208	.0308	.0363	.0359	.0392	.0413	.0478	.0398
$\hat{\beta}_{mc}^A(10, 5\sqrt{3}/24)$	0	.0213	.0754	.2052	.3253	.4365	.946	.9993	.9473	.4242
$\hat{\beta}_{mc}^A(10, \sqrt{3}/12)$	0	.0921	.0645	.1448	.2002	.2274	.2739	.1383	.0823	.0639
$\hat{\beta}_{mc}^A(10, \sqrt{3}/21)$	0	.0151	.0364	.0605	.0746	.0771	.0764	.0618	.0501	.0518

Table 5: The empirical critical values, empirical significance levels, and empirical power estimates under  $H_\epsilon^A$  for  $\epsilon = 5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21$  and  $n = 10$  at  $\alpha = .05$ .

kernel density estimates are almost symmetric for  $r = \sqrt{2}, 3/2, 2$ , with most symmetry occurring at  $r = 3/2$ , kernel density estimates are skewed left for  $r = 3, 5, 10$ , (with skewness increasing as  $r$  gets larger).

We also plot the empirical power as a function of  $r$  in Figure 16. Let  $r_A^*(\epsilon)$  be the value at which maximum Monte Carlo power estimate occurs. Then  $r_A^*(5\sqrt{3}/24) = 3$ ,  $r_A^*(\sqrt{3}/12) = 2$ , and for  $r_A^*(\sqrt{3}/21) = 3/2$ . Notice that the more severe the association the larger the value of  $r_A^*$ . Based on the analysis of the Monte Carlo power estimates, we suggest moderate  $r$  values for moderate association.

We also estimate the power using the asymptotic critical value in association alternatives for various values of  $r$ . For each  $r$  value, the level  $\alpha$  asymptotic critical value is  $\mu(r) + z_\alpha \cdot \sqrt{\nu(r)/n}$ . We estimate the empirical power as  $\hat{\beta}_n^A(r, \epsilon) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(R_j < z_\alpha)$ .

In Figure 17, we present a Monte Carlo investigation of power against  $H_{\sqrt{3}/21}^A$ ,  $H_{\sqrt{3}/12}^A$ , and  $H_{5\sqrt{3}/24}^A$  as a function of  $r$  for  $n = 10$ . The empirical significance level is  $\hat{\alpha}_A(n) := \frac{1}{N} \sum_{j=1}^N \mathbf{I}(R_j < z_\alpha | H_0)$ . Then  $\hat{\alpha}_A(10)$ , is about .05 for  $r = \sqrt{2}, 3/2, 2, 3, 5$  which have the empirical power  $\hat{\beta}_{10}^A(r, \sqrt{3}/12) \leq .35$  with maximum power at  $r = 2$ , and  $\hat{\beta}_{10}^A(r = 3, 5\sqrt{3}/24) = 1$ . So, for small sample sizes, moderate values of  $r$  are more appropriate for normal approximation, as they yield the desired significance level, and the more severe the association, the higher the power estimate.

The empirical significance levels and empirical power  $\hat{\beta}_n^S(r, \epsilon)$  values under  $H_\epsilon^A$  for  $\epsilon = 5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21$  are presented in Table 6.

Note that even for  $n = 10$ , the plots of the empirical power  $\hat{\beta}_{10}^A(r)$  resembles the curves of the asymptotic power function  $\Pi_A(r)$  in Section 3.9.

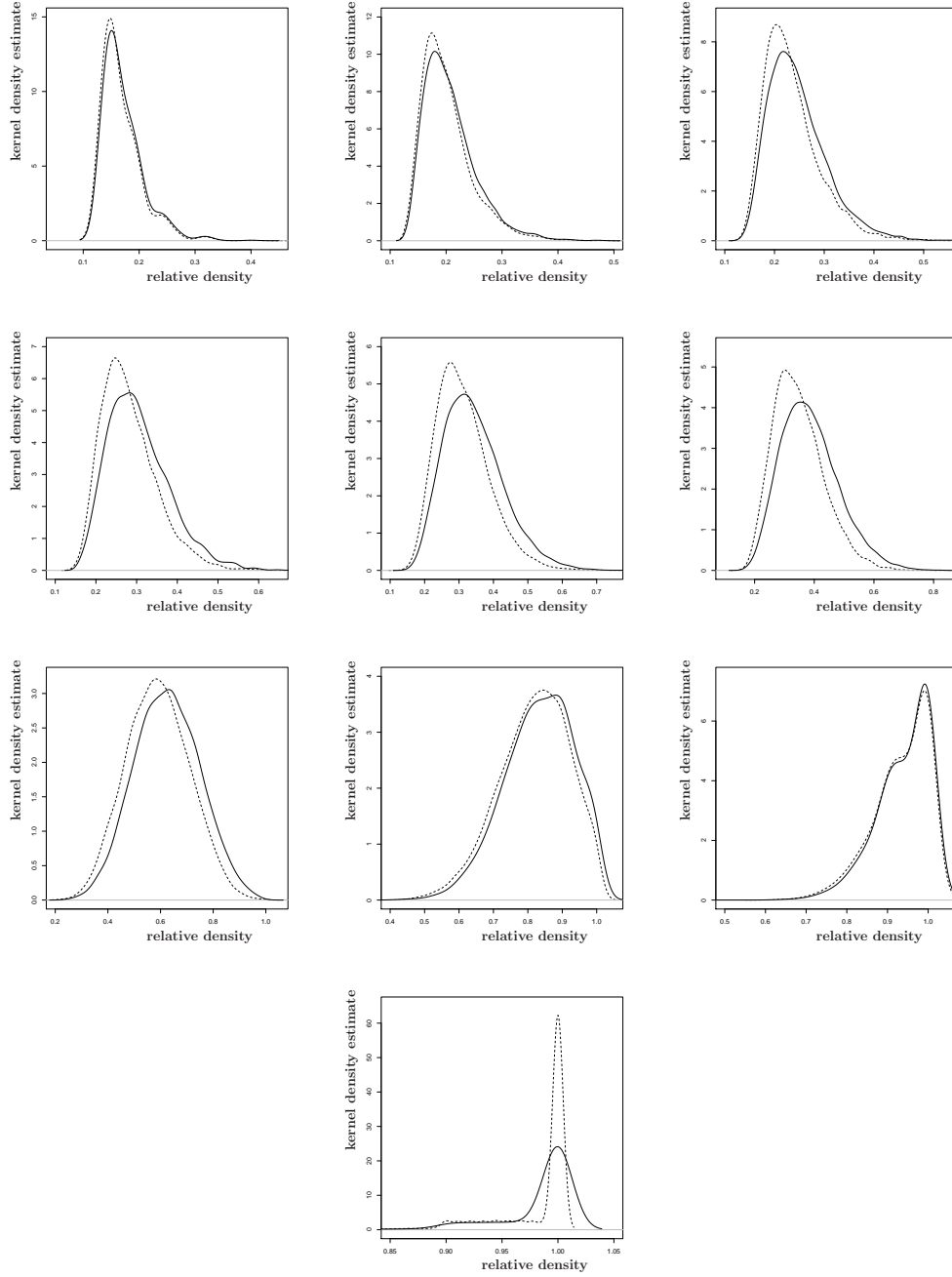


Figure 15: Kernel density estimates for the null (solid) and the association alternative  $H_{\sqrt{3}/21}^A$  (dashed) for  $r = 1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2, 2, 3, 5, 10$  (left-to-right).

$r$	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5	10
$\hat{\alpha}_A(10)$	.5318	.2426	.1869	.1031	.0673	.0559	.0656	.0627	.0771	.0955
$\hat{\beta}_{10}^A(r, 5\sqrt{3}/24)$	.6273	.3663	.3923	.4103	.4167	.5316	.9610	.9983	.9656	.5443
$\hat{\beta}_{10}^A(r, \sqrt{3}/12)$	.6300	.3537	.3583	.3190	.2698	.2919	.3433	.1825	.1429	.1261
$\hat{\beta}_{10}^A(r, \sqrt{3}/21)$	.6012	.2979	.2574	.1629	.1190	.1077	.1098	.0889	.0989	.1033

Table 6: The empirical significance level and empirical power values under  $H_{\epsilon}^A$  for  $\epsilon = 5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21$  with  $N = 10,000$ , and  $n = 10$  at  $\alpha = .05$ .

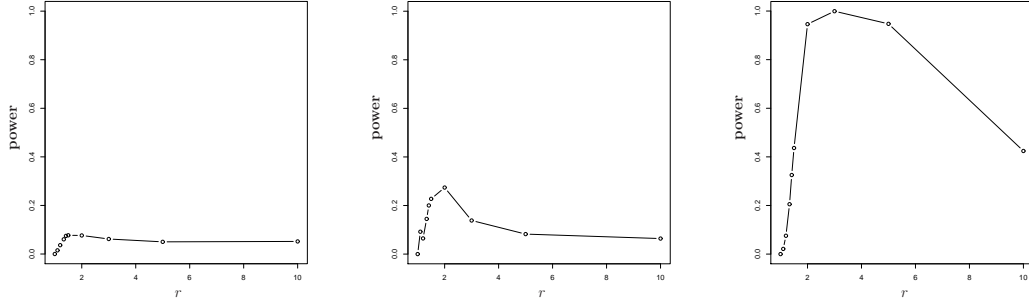


Figure 16: Monte Carlo power using the empirical critical value against association alternatives  $H^A_{\sqrt{3}/21}$  (left),  $H^A_{\sqrt{3}/12}$  (middle) and  $H^A_{5\sqrt{3}/24}$  (right) as a function of  $r$ , for  $n = 10$ .

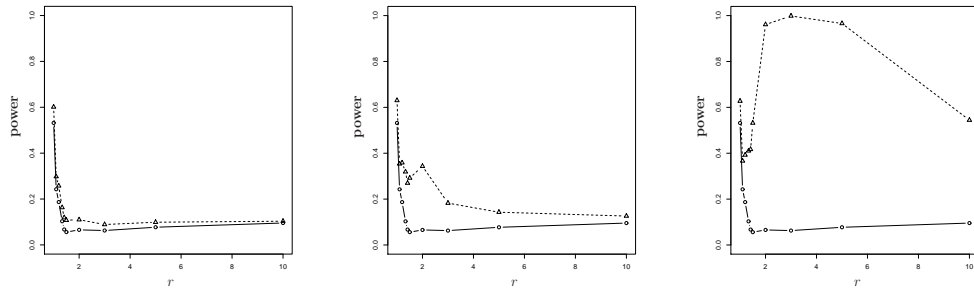


Figure 17: Monte Carlo power using the asymptotic critical value against association alternatives  $H^A_{\sqrt{3}/21}$  (left),  $H^A_{\sqrt{3}/12}$  (middle), and  $H^A_{5\sqrt{3}/24}$  (right) as a function of  $r$ , for  $n = 10$ . The circles represent the empirical significance levels while triangles represent the empirical power values.

### 3.7 Pitman Asymptotic Efficacy

Suppose that the distribution  $F$  under consideration may be indexed by a set  $\Theta \subset \mathbb{R}$  and consider  $H_0 : \theta = \theta_0$  versus  $H_a : \theta > \theta_0$ .

Pitman asymptotic efficacy (PAE) provides for an investigation of “local asymptotic power” — local around  $H_0$ . This involves the limit as  $n \rightarrow \infty$  as well as the limit as  $\epsilon \rightarrow 0$ .

Consider the comparison of test sequences  $S = \{S_n\}$  satisfying the following conditions in a neighborhood  $\theta \in [\theta_0, \theta_0 + \delta]$  of the null parameter for some  $\delta > 0$ .

**Pitman’s Conditions:**

(PC1) For some functions  $\mu_n(\theta)$  and  $\sigma_n(\theta)$ , the distribution  $F_\theta$  of  $[S_n - \mu_n(\theta)]/\sigma_n(\theta)$  converges to  $Z \sim \mathcal{N}(0, 1)$  uniformly on  $[\theta_0, \theta_0 + \delta]$ , i.e.,

$$\sup_{\theta_0 \leq \theta \leq \theta_0 + \delta} \sup_{t \in \mathbb{R}} \left| P \left( \frac{S_n - \mu_n(\theta)}{\sigma_n(\theta)} \leq t \right) - \Phi(t) \right| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

(PC2) For  $\theta \in [\theta_0, \theta_0 + \delta]$ ,  $\mu_n(\theta)$  is differentiable with  $\mu'_n(\theta_0) > 0$ ,

(PC3) For  $\theta_n = \theta_0 + O(n^{-1/2})$ ,  $\lim_{n \rightarrow \infty} \frac{\mu'_n(\theta_n)}{\mu'_n(\theta_0)} = 1$ ,

(PC4) For  $\theta_n = \theta_0 + O(n^{-1/2})$ ,  $\lim_{n \rightarrow \infty} \frac{\sigma_n(\theta_n)}{\sigma_n(\theta_0)} = 1$ .

(PC5) For some constant  $c > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{\mu'_n(\theta_0)}{\sqrt{n} \sigma_n(\theta_0)} = c,$$

Condition (PC1) is equivalent to

(PC1)’ For some functions  $\mu_n(\theta)$  and  $\sigma_n(\theta)$ , the distribution  $F_\theta$  of  $[S_n - \mu_n(\theta_n)]/\sigma_n(\theta_n)$  converges to a standard normal distribution (see Eeden (1963)).

Note that if  $\mu_n^{(k)}(\theta_0) > 0$  and  $\mu_n^{(l)}(\theta_0) = 0$ , for all  $l = 1, 2, \dots, k-1$ , then  $\mu'_n(\theta_0)$  in (PC2), (PC3), and (PC5) can be replaced by  $\mu_n^{(k)}(\theta_0) > 0$  and  $\mu'_n(\theta_n)$  in (PC3) can be replaced by  $\mu_n^{(k)}(\theta_n)$  (see Kendall and Stuart (1979)).

**Lemma 1:** (Pitman-Noether)

(i) Let  $S = \{S_n\}$  satisfy (PC1)-(PC5). Consider testing  $H_0$  by the critical regions  $S_n > u_{\alpha_n}$  with  $\alpha_n = P_{\theta_0}(S_n > u_{\alpha_n}) \rightarrow \alpha$  as  $n \rightarrow \infty$  where  $\alpha \in (0, 1)$ . For  $\beta \in (0, 1 - \alpha)$  and  $\theta_n = \theta_0 + O(n^{-1/2})$ , we have

$$\beta_n(\theta_n) = P_{\theta_n}(T_n > u_{\alpha_n}) \rightarrow \beta \text{ iff } c \sqrt{n}(\theta_n - \theta_0) \rightarrow \Phi^{-1}(1 - \alpha) - \Phi^{-1}(\beta).$$

(ii) Let  $S = \{S_n\}$  and  $Q = \{Q_n\}$  each satisfy (PC1)-(PC5). Then the asymptotic relative efficiency of  $S$  relative to  $Q$  is given by  $ARE(S, Q) = (c_S/c_Q)^2$ .

Thus, to evaluate  $ARE(S, Q)$  under the conditions (PC1)-(PC5), we need only calculate the quantities  $c_S$  and  $c_Q$ , where

$$c_S = \lim_{n \rightarrow \infty} \frac{\mu'_{S_n}(\theta_0)}{\sqrt{n} \cdot \sigma_{S_n}(\theta_0)} \text{ and } c_Q = \lim_{n \rightarrow \infty} \frac{\mu'_{Q_n}(\theta_0)}{\sqrt{n} \cdot \sigma_{Q_n}(\theta_0)}$$

$PAE(S) = c_S^2$  is called the *Pitman Asymptotic Efficacy* (PAE) of the test based on  $S_n$ . Using similar notation and terminology for  $Q_n$ ,

$$ARE(S, Q) = \frac{PAE(S)}{PAE(Q)}.$$

For segregation or association alternatives the PAE of  $\rho_n(r)$  is given by  $PAE(r) = \frac{(\mu^{(k)}(r, \epsilon=0))^2}{\nu(r)}$  where  $k$  is the minimum order of the derivative with respect to  $\epsilon$  for which  $\mu^{(k)}(r, \epsilon=0) \neq 0$ . That is,  $\mu^{(k)}(r, \epsilon=0) \neq 0$  but  $\mu^{(l)}(r, \epsilon=0) = 0$  for  $l = 1, 2, \dots, k-1$ .

### 3.7.1 Pitman Asymptotic Efficacy Under Segregation Alternatives

Consider the test sequences  $\rho(r) = \{\rho_n(r)\}$  for sufficiently small  $\epsilon > 0$  and  $r \in [1, \sqrt{3}/(2\epsilon)]$ .

In the PAE framework above,  $\theta = \epsilon$  and  $\theta_0 = 0$ . Suppose,  $\mu_n(\epsilon) = E_\epsilon^S[\rho_n(r)] = \mu_S(r, \epsilon)$ . For  $\epsilon \in [0, \sqrt{3}/8)$ ,

$$\mu_S(r, \epsilon) = \sum_{j=1}^5 \varpi_{1,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$$

with the corresponding intervals  $\mathcal{I}_1 = [1, 3/2 - \sqrt{3}\epsilon)$ ,  $\mathcal{I}_2 = [3/2 - \sqrt{3}\epsilon, 3/2)$ ,  $\mathcal{I}_3 = [3/2, 2 - 4\epsilon/\sqrt{3})$ ,  $\mathcal{I}_4 = [2 - 4\epsilon/\sqrt{3}, 2)$ ,  $\mathcal{I}_5 = [2, \sqrt{3}/(2\epsilon))$ . See Appendix 2 for the explicit form of  $\mu(r, \epsilon)$  and Appendix 3 for derivation. Notice that as  $\epsilon \rightarrow 0$ , only  $\mathcal{I}_1 = [1, 3/2 - \sqrt{3}\epsilon)$ ,  $\mathcal{I}_3 = [3/2, 2 - 4\epsilon/\sqrt{3})$ ,  $\mathcal{I}_5 = [2, \sqrt{3}/(2\epsilon))$  do not vanish, so we only keep the components of  $\mu_S(r, \epsilon)$  on these intervals.

Furthermore,  $\sigma_S^2(n, \epsilon) = \mathbf{Var}_\epsilon^S(\rho_n(r)) = \frac{1}{2n(n-1)} \mathbf{Var}_\epsilon^S[h_{12}] + \frac{(n-2)}{n(n-1)} \nu_S(r, \epsilon) = \mathbf{Cov}_\epsilon^S[h_{12}, h_{13}]$ . The explicit forms of  $\mathbf{Var}_\epsilon^S[h_{12}]$  and  $\mathbf{Cov}_\epsilon^S[h_{12}, h_{13}]$  are not calculated, since we only need  $\lim_{n \rightarrow \infty} \sigma_n^2(\epsilon = 0) = \nu(r)$  which is given in Equation 9.

Notice that  $\mathbf{E}_\epsilon^S[h_{12}]^3 \leq 8 < \infty$  and  $\mathbf{E}_\epsilon^S[h_{12} h_{13}] - \mathbf{E}_\epsilon^S[h_{12}]^2 = \mathbf{Cov}_\epsilon^S[h_{12}, h_{13}] > 0$  then by Callaert and Janssen (1978)

$$\sup_{t \in \mathbb{R}} \left| P_\epsilon \left( \sqrt{n} \frac{(\rho_n(r) - \mu_S(r, \epsilon))}{\sqrt{\nu_S(r, \epsilon)}} \leq t \right) - \Phi(t) \right| \leq C \mathbf{E}_\epsilon^S |h_{12}|^3 [\nu_S(r, \epsilon)]^{-\frac{3}{2}} n^{-\frac{1}{2}}$$

where  $C$  is an absolute constant and  $\Phi(\cdot)$  is the standard normal distribution function. Then (PC1) follows for each  $r \in [1, \sqrt{3}/(2\epsilon))$  and  $\epsilon \in [0, \sqrt{3}/4)$ .

Differentiating  $\mu_S(r, \epsilon)$  with respect to  $\epsilon$  yields

$$\begin{aligned} \mu'_S(r, \epsilon) = & \varpi'_{1,1}(r, \epsilon) \mathbf{I}(r \in [1, 3/2 - \sqrt{3}\epsilon)) + \varpi'_{1,3}(r, \epsilon) \mathbf{I}(r \in [3/2, 2 - 4\epsilon/\sqrt{3})) \\ & + \varpi'_{1,5}(r, \epsilon) \mathbf{I}(r \in [2, \sqrt{3}/(2\epsilon))) \end{aligned}$$

where

$$\begin{aligned} \varpi'_{1,1}(r, \epsilon) &= \frac{2\epsilon(144\epsilon^2(r^2 - 1) + 36 - 37r^2)}{27(2\epsilon - 1)^3(2\epsilon + 1)^3} \\ \varpi'_{1,3}(r, \epsilon) &= \left[ 2\sqrt{3} \left( (2r - 3)64\epsilon^3 + (7r^2 + r^4 - 24r + 20)16\sqrt{3}\epsilon^2 + (r - 3)48\epsilon + 3\sqrt{3}r^4 + 96\sqrt{3}r \right. \right. \\ & \quad \left. \left. - 36\sqrt{3} - 60\sqrt{3}r^2 \right) \epsilon \right] / \left[ 9(2\epsilon + 1)^3(2\epsilon - 1)^3r^2 \right] \\ \varpi'_{1,5}(r, \epsilon) &= \frac{8\sqrt{3}\epsilon(48\epsilon^3 + (3r^4 + 3r^2 - 20)4\sqrt{3}\epsilon^2 + 36\epsilon + 9\sqrt{3} - 9\sqrt{3}r^2)}{27r^2(2\epsilon + 1)^3(2\epsilon - 1)^3}. \end{aligned}$$

Hence,  $\mu'_S(r, \epsilon = 0) = 0$ , so we need higher order derivatives for (PC2). A detailed discussion is available in Kendall and Stuart (1979).

Differentiating  $\mu'_S(r, \epsilon)$  with respect to  $\epsilon$  yields

$$\begin{aligned} \mu''_S(r, \epsilon) = & \varpi''_{1,1}(r, \epsilon) \mathbf{I}(r \in [1, 3/2 - \sqrt{3}\epsilon)) + \varpi''_{1,3}(r, \epsilon) \mathbf{I}(r \in [3/2, 2 - 4\epsilon/\sqrt{3})) \\ & + \varpi''_{1,5}(r, \epsilon) \mathbf{I}(r \in [2, \sqrt{3}/(2\epsilon))) \end{aligned}$$

where

$$\begin{aligned}\varpi''_{1,1}(r, \epsilon) &= -\frac{2(r^2 - 1)1728\epsilon^4 + (72 - 77r^2)4\epsilon^2 + 36 - 37r^2}{27(4\epsilon^2 - 1)^4} \\ \varpi''_{1,3}(r, \epsilon) &= -2 \left[ (2r - 3)512\sqrt{3}\epsilon^5 + (20 + r^4 + 7r^2 - 24r)576\epsilon^4 + (2r - 3)1024\sqrt{3}\epsilon^3 + (20 - 108r^2 \right. \\ &\quad \left. + 96r + 9r^4)36\epsilon^2 + (-3 + 2r)96\sqrt{3}\epsilon - 108 + 9r^4 - 180r^2 + 288r \right] / \left[ 9r^2(2\epsilon + 1)^4(2\epsilon - 1)^4 \right] \\ \varpi''_{1,5}(r, \epsilon) &= -8 \left[ 128\sqrt{3}\epsilon^5 + (-20 + 3r^4 + 3r^2)48\epsilon^4 + 256\sqrt{3}\epsilon^3 + (-5 - 12r^2 + 3r^4)12\epsilon^2 \right. \\ &\quad \left. + 24\epsilon\sqrt{3} + 9 - 9r^2 \right] / \left[ 9r^2(2\epsilon + 1)^4(2\epsilon - 1)^4 \right].\end{aligned}$$

Thus,

$$\mu''_S(r, \epsilon = 0) = \begin{cases} -\frac{8}{3} + \frac{74}{27}r^2 & \text{for } r \in [1, 3/2) \\ -2 \frac{(r^2 - 4)r + 2)(r^2 + 4r - 6)}{r^2} & \text{for } r \in [3/2, 2) \\ -\frac{8(1 - r^2)}{r^2} & \text{for } r \in [2, \sqrt{3}/(2\epsilon)). \end{cases} \quad (12)$$

Observe that  $\mu''_S(r, \epsilon = 0) > 0$  for all  $r \in [1, \sqrt{3}/(2\epsilon))$ , so (PC2) holds with the second derivative. (PC3) in the second derivative form follows from continuity of  $\mu''_S(r, \epsilon)$  in  $\epsilon$  and (PC4) follows from continuity of  $\sigma_n^2(r, \epsilon)$  in  $\epsilon$ .

Next, we find  $c_S(\rho(r)) = \lim_{n \rightarrow \infty} \frac{\mu''_S(r, \epsilon = 0)}{\sqrt{n} \sigma_n(r, \epsilon = 0)} = \frac{\mu''_S(r, \epsilon = 0)}{\sqrt{\nu(r)}}$ , where numerator is given in Equation 12 and denominator is given in Equation 9. We can easily see that  $c_S(\rho(r)) > 0$ , since  $c_S(\rho(r))$  is increasing in  $r$  and  $c_S(\rho(r = 1)) > 0$ . Then (PC5) follows. So under segregation alternatives  $H_\epsilon^S$ , the PAE of  $\rho_n(r)$  is given by

$$\text{PAE}^S(r) = c_S^2(\rho(r)) = \frac{(\mu''_S(r, \epsilon = 0))^2}{\nu(r)}.$$

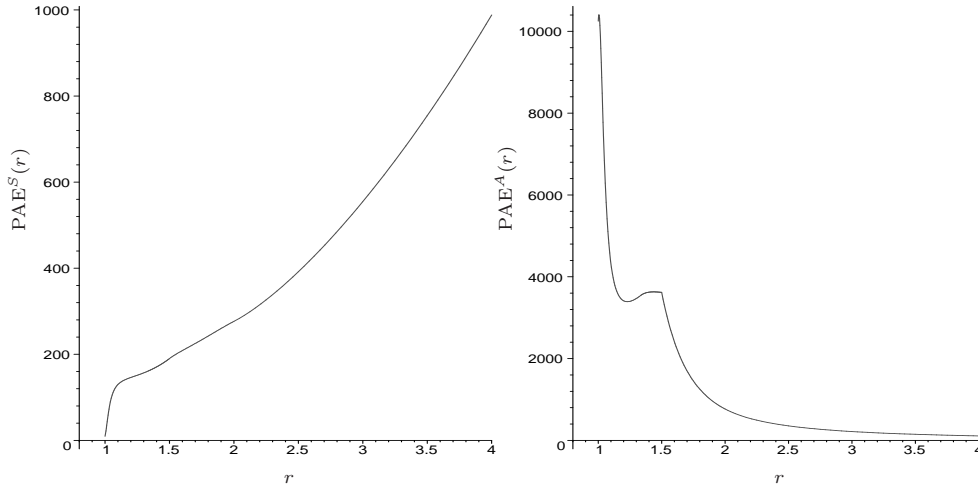


Figure 18: Pitman asymptotic efficacy against segregation (left) and against association (right) as a function of  $r$ .

In Figure 18 (left), we present the PAE as a function of  $r$  for segregation. Notice that  $\text{PAE}^S(r = 1) = 160/7 \approx 22.8571$ ,  $\lim_{r \rightarrow \infty} \text{PAE}^S(r) = \infty$ . Based on the PAE analysis, we suggest, for large  $n$  and small  $\epsilon$ , choosing  $r$  large for testing against segregation. However, for small and moderate values of  $n$ , normal approximation is not appropriate due to the skewness in the density of  $\rho_n(r)$ , Therefore, for small  $n$ , we suggest moderate  $r$  values.

PAE analysis is local (around  $\epsilon = 0$ ) and for arbitrarily large  $n$ . The comparison would hold in general provided that  $\mu(r, \epsilon)$  is convex in  $\epsilon$  for all  $\epsilon \in [0, \sqrt{3}/3)$ . As an alternative, we fix an  $\epsilon$  and then compare the asymptotic behaviour of  $\rho_n(r)$  with Hodges-Lehmann asymptotic efficacy in Section 3.8.1.

### 3.7.2 Pitman Asymptotic Efficacy Under Association Alternatives

Consider the test sequences  $\rho(r) = \{\rho_n(r)\}$  for sufficiently small  $\epsilon > 0$  and  $r \in [1, \infty)$ .

In the PAE framework above,  $\theta = \epsilon$  and  $\theta_0 = 0$ . Suppose,  $\mu_n(\epsilon) = \mathbf{E}_\epsilon[\rho_n(r)] = \mu_A(r, \epsilon)$ . For  $\epsilon \in [0, (7\sqrt{3} - 3\sqrt{15})/12 \approx .042)$ ,

$$\mu_A(r, \epsilon) = \sum_{j=1}^6 \varpi_{1,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$$

with the corresponding intervals  $\mathcal{I}_1 = [1, (1 + 2\sqrt{3}\epsilon) / (1 - \sqrt{3}\epsilon)]$ ,  $\mathcal{I}_2 = [(1 + 2\sqrt{3}\epsilon) / (1 - \sqrt{3}\epsilon), 4(1 - \sqrt{3}\epsilon)/3]$ ,  $\mathcal{I}_3 = [4(1 - \sqrt{3}\epsilon)/3, 4(1 + 2\sqrt{3}\epsilon)/3]$ ,  $\mathcal{I}_4 = [4(1 + 2\sqrt{3}\epsilon)/3, 3/(2(1 - \sqrt{3}\epsilon))]$ ,  $\mathcal{I}_5 = [3/(2(1 - \sqrt{3}\epsilon)), 2]$  and  $\mathcal{I}_6 = [2, \infty)$ . Notice that as  $\epsilon \rightarrow 0$ , only  $\mathcal{I}_j$  for  $j = 2, 4, 5, 6$  do not vanish, so we only keep the components of  $\mu_A(r, \epsilon)$  on these intervals. See Section ?? for the explicit form of  $\mu_A(r, \epsilon)$  and Section ?? for derivation.

Furthermore,  $\sigma_n^2(\epsilon) = \mathbf{Var}_\epsilon^A(\rho_n(r)) = \frac{1}{2n(n-1)} \mathbf{Var}_\epsilon^A[h_{12}] + \frac{(n-2)}{n(n-1)} \mathbf{Cov}_\epsilon^A[h_{12}, h_{13}]$  whose explicit form is not calculated, since we only need  $\lim_{n \rightarrow \infty} \sqrt{n} \sigma_n(\epsilon = 0) = \nu(r)$  which is given Equation 9.

(PC1) follows for each  $r \in [1, \infty)$  and  $\epsilon \in [0, \sqrt{3}/3)$  as in the segregation case.

Differentiating  $\mu_A(r, \epsilon)$  with respect to  $\epsilon$ , then we get

$$\begin{aligned} \mu'_A(r, \epsilon) = & \varpi'_{1,2}(r, \epsilon) \mathbf{I}(r \in [1, 4/3)) + \varpi'_{1,4}(r, \epsilon) \mathbf{I}(r \in [4/3, 3/2)) \\ & + \varpi'_{1,5}(r, \epsilon) \mathbf{I}(r \in [3/2, 2)) + \varpi'_{1,6}(r, \epsilon) \mathbf{I}(r \in [2, \infty)) \end{aligned}$$

where

$$\begin{aligned} \varpi'_{1,2}(r, \epsilon) = & -2 \left[ \sqrt{3} \left( -1152 r^4 \epsilon^3 + 720 \sqrt{3} r^4 \epsilon^2 - 288 r^4 \epsilon + 11 \sqrt{3} r^4 + 2592 \sqrt{3} r^2 \epsilon^2 - 10368 \sqrt{3} r \epsilon^2 \right. \right. \\ & \left. \left. + 432 \sqrt{3} r^2 + 6480 \sqrt{3} \epsilon^2 - 864 \sqrt{3} r + 432 \sqrt{3} \right) \epsilon \right] / \left[ \left( -6 \epsilon + \sqrt{3} \right)^3 \left( 6 \epsilon + \sqrt{3} \right)^3 r^2 \right], \\ \varpi'_{1,4}(r, \epsilon) = & -2 \left[ \sqrt{3} \left( -1152 r^4 \epsilon^3 + 720 \sqrt{3} r^4 \epsilon^2 - 288 r^4 \epsilon + 11 \sqrt{3} r^4 - 1296 \sqrt{3} r^2 \epsilon^2 + 108 \sqrt{3} r^2 \right. \right. \\ & \left. \left. - 2160 \sqrt{3} \epsilon^2 - 144 \sqrt{3} \right) \epsilon \right] / \left[ \left( -6 \epsilon + \sqrt{3} \right)^3 \left( 6 \epsilon + \sqrt{3} \right)^3 r^2 \right], \\ \varpi'_{1,5}(r, \epsilon) = & \frac{2 \epsilon (3 r^4 - 72 r^2 - 240 \epsilon^2 + 192 r - 124)}{r^2 (12 \epsilon^2 - 1)^3}, \\ \varpi'_{1,6}(r, \epsilon) = & -\frac{40 \epsilon}{r^2 (12 \epsilon^2 - 1)^2}. \end{aligned}$$

Hence  $\mu'_A(r, \epsilon = 0) = 0$ , so we differentiate  $\mu'_A(r, \epsilon)$  with respect to  $\epsilon$  and get

$$\begin{aligned} \mu''_A(r, \epsilon) = & \varpi''_{1,2}(r, \epsilon) \mathbf{I}(r \in [1, 4/3)) + \varpi''_{1,4}(r, \epsilon) \mathbf{I}(r \in [4/3, 3/2)) \\ & + \varpi''_{1,5}(r, \epsilon) \mathbf{I}(r \in [3/2, 2)) + \varpi''_{1,6}(r, \epsilon) \mathbf{I}(r \in [2, \infty)) \end{aligned}$$



where

$$\begin{aligned}
\varpi''_{1,2}(r, \epsilon) &= -6 \left[ \sqrt{3} \left( -27648 r^4 \epsilon^5 + 25920 \sqrt{3} r^4 \epsilon^4 - 18432 r^4 \epsilon^3 + 2820 \sqrt{3} r^4 \epsilon^2 + 93312 \sqrt{3} r^2 \epsilon^4 \right. \right. \\
&\quad \left. \left. - 576 r^4 \epsilon - 373248 \sqrt{3} r \epsilon^4 + 11 \sqrt{3} r^4 + 33696 \sqrt{3} r^2 \epsilon^2 + 233280 \sqrt{3} \epsilon^4 - 82944 \sqrt{3} r \epsilon^2 \right. \right. \\
&\quad \left. \left. + 432 \sqrt{3} r^2 + 45360 \sqrt{3} \epsilon^2 - 864 \sqrt{3} r + 432 \sqrt{3} \right) \right] / \left[ \left( 6 \epsilon + \sqrt{3} \right)^4 \left( -6 \epsilon + \sqrt{3} \right)^4 r^2 \right], \\
\varpi''_{1,4}(r, \epsilon) &= -6 \left[ \sqrt{3} \left( -27648 r^4 \epsilon^5 + 25920 \sqrt{3} r^4 \epsilon^4 - 18432 r^4 \epsilon^3 + 2820 \sqrt{3} r^4 \epsilon^2 - 46656 \sqrt{3} r^2 \epsilon^4 \right. \right. \\
&\quad \left. \left. - 576 r^4 \epsilon + 11 \sqrt{3} r^4 + 2592 \sqrt{3} r^2 \epsilon^2 - 77760 \sqrt{3} \epsilon^4 + 108 \sqrt{3} r^2 - 15120 \sqrt{3} \epsilon^2 - 144 \sqrt{3} \right) \right] \\
&\quad / \left[ \left( 6 \epsilon + \sqrt{3} \right)^4 \left( -6 \epsilon + \sqrt{3} \right)^4 r^2 \right], \\
\varpi''_{1,5}(r, \epsilon) &= -\frac{2(180 r^4 \epsilon^2 + 3 r^4 - 4320 r^2 \epsilon^2 - 8640 \epsilon^4 + 11520 r \epsilon^2 - 72 r^2 - 8160 \epsilon^2 + 192 r - 124)}{r^2(12 \epsilon^2 - 1)^4}, \\
\varpi''_{1,6}(r, \epsilon) &= \frac{40(36 \epsilon^2 + 1)}{r^2(12 \epsilon^2 - 1)^3}.
\end{aligned}$$

Thus,

$$\mu''_A(r, \epsilon = 0) = \begin{cases} -\frac{22}{9} r^2 + 192 r^{-1} - 96 r^{-2} - 96 & \text{for } r \in [1, 4/3) \\ -\frac{22}{9} r^2 + 32 r^{-2} - 24 & \text{for } r \in [4/3, 3/2) \\ -6 r^2 - 384 r^{-1} + 248 r^{-2} + 144 & \text{for } r \in [3/2, 2) \\ -40 r^{-2} & \text{for } r \in [2, \infty). \end{cases} \quad (13)$$

Note that  $\mu''_A(r, \epsilon = 0) > 0$  for all  $r \in [1, \infty)$ , so (PC2) follows with the second derivative. (PC3) and (PC4) follow from continuity of  $\mu''_A(r, \epsilon)$  and  $\sigma_n^2(r, \epsilon)$  in  $\epsilon$ .

Next, we find  $c_A(\rho(r)) = \lim_{n \rightarrow \infty} \frac{\mu''_A(r, \epsilon = 0)}{\sqrt{n} \sigma_n(r, \epsilon = 0)} = \frac{\mu''_A(r, 0)}{\sqrt{\nu(r)}}$ , by substituting the numerator from Equation 13 and denominator from Equation 9. We can easily see that  $c_A(\rho(r)) < 0$ , for all  $r \geq 1$ . Then (PC5) holds, so under association alternatives  $H_\epsilon^A$ , the PAE of  $\rho_n(r)$  is

$$\text{PAE}^A(r) = c_A^2(\rho(r)) = \frac{(\mu''_A(r, \epsilon = 0))^2}{\nu(r)}.$$

In Figure 18 (right), we present the PAE as a function of  $r$  for association. Notice that  $\text{PAE}^A(r = 1) = 174240/17 \approx 10249.4118$ ,  $\lim_{r \rightarrow \infty} \text{PAE}^A(r) = 0$ ,  $\text{argsup}_{r \in [1, \infty)} \text{PAE}^A(r) \approx 1.006$  with supremum  $\approx 10399.7726$ .  $\text{PAE}^A(r)$  has also a local supremum at  $r_l \approx 1.4356$  with local supremum  $\approx 3630.8932$ . Based on the Pitman asymptotic efficacy analysis, we suggest, for large  $n$  and small  $\epsilon$ , choosing  $r$  small for testing against association. However, for small and moderate values of  $n$  normal approximation is not appropriate due to the skewness in the density of  $\rho_n(r)$ . Therefore, for small  $n$ , we suggest moderate  $r$  values.

We also calculate Hodges-Lehmann asymptotic efficacy for fixed alternatives in Section 3.8.2.

### 3.8 Hodges-Lehmann Asymptotic Efficacy

Unlike PAE, HLAE does not involve the limit as  $\epsilon \rightarrow 0$ . Since this requires the mean and, especially, the asymptotic variance of  $\rho_n(r)$  under an alternative, we investigate HLAE for specific values of  $\epsilon$ . See Appendix 4 for a sample derivation of  $\mu_S(r, \epsilon)$  and  $\nu_S(r, \epsilon)$ .

#### 3.8.1 Hodges-Lehmann Asymptotic Efficacy Under Segregation Alternatives

In the HLAE framework,  $\theta = \mathbf{E}_\epsilon^S[\rho_n(r)] = \mu_S(r, \epsilon)$  and  $\theta_0 = \mu(r)$ . Then testing  $H_0 : \epsilon = 0$  versus  $H_\epsilon^S : \epsilon > 0$  is equivalent to  $H_0 : \mathbf{E}[\rho_n] = \mu(r)$  versus  $H_\epsilon^S : \mathbf{E}_\epsilon^S[\rho_n] = \mu_S(r, \epsilon) > \mu(r)$ . Let  $\delta = \frac{\mu_S(r, \epsilon) - \mu(r)}{\sqrt{\nu_S(r, \epsilon)}}$  and  $\tilde{R} = \frac{\rho_n(r) - \mu(r)}{\sqrt{\nu_S(r, \epsilon)}}$ , then  $\tilde{R} \xrightarrow{L} \mathcal{N}(\delta, 1)$ .

Then HLAE of  $\rho_n(r)$  is given by

$$\text{HLAE}^S(r, \epsilon) := \frac{(\mu_S(r, \epsilon) - \mu(r))^2}{\nu_S(r, \epsilon)}.$$

We calculate HLAE of  $\rho_n(r)$  under  $H_\epsilon^S$  for  $\epsilon = \sqrt{3}/8$ ,  $\epsilon = \sqrt{3}/4$ , and  $\epsilon = 2\sqrt{3}/7$ .

With  $\epsilon = \sqrt{3}/8$ ,  $\rho_n(r)$  is non-degenerate for  $r \in [1, 4)$

$$\mu_S(r, \sqrt{3}/8) = \begin{cases} \frac{2287}{9126} r^2 - \frac{1}{13} & \text{for } r \in [1, 9/8) \\ -\frac{5905 r^4 - 36864 r^3 + 62910 r^2 - 46656 r + 13122}{9126 r^2} & \text{for } r \in [9/8, 3/2) \\ \frac{61 r^4 - 768 r^3 + 3494 r^2 - 5120 r + 2466}{338 r^2} & \text{for } r \in [3/2, 2) \\ -\frac{3 r^4 - 422 r^2 + 606}{338 r^2} & \text{for } r \in [2, 3) \\ \frac{3 r^4 - 48 r^3 + 530 r^2 - 768}{338 r^2} & \text{for } r \in [2, 4] \end{cases}$$

and

$$\nu_S(r, \sqrt{3}/8) = \sum_{j=1}^{12} \nu_j(r, \sqrt{3}/8) \mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned} \nu_1(r, \sqrt{3}/8) &= \left[ 9959911 r^{10} - 46006272 r^9 - 430526 r^8 + 258785280 r^7 - 385799609 r^6 + 162699264 r^5 \right. \\ &\quad \left. - 83976048 r^4 + 201277440 r^3 - 129392640 r^2 + 12939264 \right] / [104104845 r^4], \\ \nu_2(r, \sqrt{3}/8) &= \left[ 9959911 r^{10} - 46006272 r^9 - 430526 r^8 + 258785280 r^7 - 415110891 r^6 + 272331072 r^5 \right. \\ &\quad \left. - 158725008 r^4 - 16174080 r^3 + 315394560 r^2 - 310542336 r + 90574848 \right] / [104104845 r^4], \\ \nu_3(r, \sqrt{3}/8) &= \left[ 3144167 r^{12} + 15335424 r^{11} - 378655166 r^{10} + 2750459904 r^9 - 11800111467 r^8 \right. \\ &\quad \left. + 31878202752 r^7 - 54792387144 r^6 + 60339341664 r^5 - 42745183272 r^4 + 19903426272 r^3 \right. \\ &\quad \left. - 6790168926 r^2 + 1989715104 r - 373071582 \right] / [104104845 r^6], \end{aligned}$$

$$\begin{aligned}
\nu_4 \left( r, \sqrt{3}/8 \right) &= - \left[ 8177689 r^{12} - 54153216 r^{11} + 320428478 r^{10} - 2459326464 r^9 + 11854698987 r^8 \right. \\
&\quad - 32751603072 r^7 + 55010737224 r^6 - 59029241184 r^5 + 42131073672 r^4 - 20886001632 r^3 \\
&\quad \left. + 7379714142 r^2 - 1694942496 r + 170415414 \right] / \left[ 104104845 r^6 \right], \\
\nu_5 \left( r, \sqrt{3}/8 \right) &= - \left[ 8177689 r^{12} - 54153216 r^{11} + 320428478 r^{10} - 2459326464 r^9 + 12509010411 r^8 \right. \\
&\quad - 37904305536 r^7 + 71918042184 r^6 - 88617024864 r^5 + 71256548232 r^4 - 36176875776 r^3 \\
&\quad \left. + 10724592861 r^2 - 1694942496 r + 170415414 \right] / \left[ 104104845 r^6 \right], \\
\nu_6 \left( r, \sqrt{3}/8 \right) &= - \left[ 2718937 r^{12} - 39596544 r^{11} + 434455742 r^{10} - 3154811904 r^9 + 14086429683 r^8 \right. \\
&\quad - 39680803584 r^7 + 72881433288 r^6 - 88893062496 r^5 + 71547681672 r^4 \\
&\quad \left. - 36487418112 r^3 + 10828106973 r^2 - 1694942496 r + 170415414 \right] / \left[ 104104845 r^6 \right], \\
\nu_7 \left( r, \sqrt{3}/8 \right) &= - \left[ 1027 r^{12} - 19968 r^{11} + 295626 r^{10} - 3265792 r^9 + 23210081 r^8 \right. \\
&\quad - 103077696 r^7 + 289042360 r^6 - 511170304 r^5 + 553668600 r^4 - 343186304 r^3 \\
&\quad \left. + 109133095 r^2 - 20431008 r + 5845554 \right] / \left[ 428415 r^6 \right], \\
\nu_8 \left( r, \sqrt{3}/8 \right) &= - \left[ 637 r^{12} - 19968 r^{11} + 299370 r^{10} - 3265792 r^9 + 23199551 r^8 \right. \\
&\quad - 103077696 r^7 + 289042360 r^6 - 511170304 r^5 + 553700190 r^4 - 343186304 r^3 \\
&\quad \left. + 109133095 r^2 - 20431008 r + 5788692 \right] / \left[ 428415 r^6 \right], \\
\nu_9 \left( r, \sqrt{3}/8 \right) &= - \left[ 637 r^{12} - 19968 r^{11} + 299370 r^{10} - 3265792 r^9 + 24051519 r^8 \right. \\
&\quad - 112023360 r^7 + 328179640 r^6 - 602490624 r^5 + 673558110 r^4 - 427086848 r^3 \\
&\quad \left. + 133604087 r^2 - 20431008 r + 5788692 \right] / \left[ 428415 r^6 \right], \\
\nu_{10} \left( r, \sqrt{3}/8 \right) &= \left[ 130 r^{12} - 2496 r^{11} + 22134 r^{10} - 122720 r^9 + 452225 r^8 - 1010880 r^7 + 1075400 r^6 \right. \\
&\quad \left. + 26624 r^5 - 1993566 r^4 + 5324800 r^3 - 5083895 r^2 + 303264 r - 37908 \right] / \left[ 428415 r^6 \right], \\
\nu_{11} \left( r, \sqrt{3}/8 \right) &= - \left[ 330 r^8 - 8896 r^7 + 85445 r^6 - 342624 r^5 + 332000 r^4 + 1148560 r^3 \right. \\
&\quad \left. - 1180986 r^2 - 5324800 r + 6678947 \right] / \left[ 428415 r^4 \right], \\
\nu_{12} \left( r, \sqrt{3}/8 \right) &= - \frac{(330 r^5 - 4936 r^4 + 12453 r^3 + 47388 r^2 - 12992 r - 128256)(r - 4)^3}{428415 r^4},
\end{aligned}$$

and the corresponding intervals are  $\mathcal{I}_1 = [1, 12/11)$ ,  $\mathcal{I}_2 = [12/11, 9/8)$ ,  $\mathcal{I}_3 = [9/8, \sqrt{6}/2)$ ,  $\mathcal{I}_4 = [\sqrt{6}/2, 21/16)$ ,  $\mathcal{I}_5 = [21/16, 4/3)$ ,  $\mathcal{I}_6 = [4/3, 3/2)$ ,  $\mathcal{I}_7 = [3/2, \sqrt{3})$ ,  $\mathcal{I}_8 = [\sqrt{3}, 7/4)$ ,  $\mathcal{I}_9 = [7/4, 2)$ ,  $\mathcal{I}_{10} = [2, 3)$ ,  $\mathcal{I}_{11} = [3, 7/2)$ ,  $\mathcal{I}_{12} = [7/2, 4)$ . See Section ?? for derivation and Figure 20 for the graph of  $\mu(r, \sqrt{3}/8)$  and  $\nu(r, \sqrt{3}/8)$ .

Then we get  $\text{HLAE}^S(r, \sqrt{3}/8) = \frac{(\mu_S(r, \sqrt{3}/8) - \mu(r))^2}{\nu_S(r, \sqrt{3}/8)}$  by substituting the relevant terms. See Figure 19.

With  $\epsilon = \sqrt{3}/4$ ,  $r \in [1, 2)$

$$\mu_S \left( r, \sqrt{3}/4 \right) = \begin{cases} -\frac{67}{54} r^2 + \frac{40}{9} r - 3 & \text{for } r \in [1, 3/2) \\ \frac{7 r^4 - 48 r^3 + 122 r^2 - 128 r + 48}{2 r^2} & \text{for } r \in [3/2, 2) \end{cases}$$

and

$$\nu_S \left( r, \sqrt{3}/4 \right) = \sum_{j=1}^5 \nu_j \left( r, \sqrt{3}/4 \right) \mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned}
\nu_1 \left( r, \sqrt{3}/4 \right) &= - \left[ 14285 r^7 - 28224 r^6 - 233266 r^5 + 1106688 r^4 - 2021199 r^3 + 1876608 r^2 \right. \\
&\quad \left. - 880794 r + 165888 \right] / \left[ 3645 r \right], \\
\nu_2 \left( r, \sqrt{3}/4 \right) &= - \left[ 14285 r^{10} - 28224 r^9 - 233266 r^8 + 1106688 r^7 - 1234767 r^6 - 3431808 r^5 \right. \\
&\quad \left. + 14049126 r^4 - 22228992 r^3 + 18895680 r^2 - 8503056 r + 1594323 \right] / \left[ 3645 r^4 \right], \\
\nu_3 \left( r, \sqrt{3}/4 \right) &= - \left[ 14285 r^{10} - 28224 r^9 - 233266 r^8 + 1106688 r^7 - 2545713 r^6 + 5903280 r^5 \right. \\
&\quad \left. - 13456044 r^4 + 20636208 r^3 - 18305190 r^2 + 8503056 r - 1594323 \right] / \left[ 3645 r^4 \right], \\
\nu_4 \left( r, \sqrt{3}/4 \right) &= \left[ 104920 r^8 - 111072 r^7 + 1992132 r^6 - 15844032 r^5 + 50174640 r^4 + 6377292 \right. \\
&\quad \left. - 34012224 r + 73220760 r^2 - 81881280 r^3 + 1909 r^{10} - 27072 r^9 \right] / \left[ 14580 r^4 \right], \\
\nu_5 \left( r, \sqrt{3}/4 \right) &= - \left[ -1187904 r^5 + 1331492 r^6 + 433304 r^2 + 611163 r^{10} - 850240 r^9 - 198144 r \right. \\
&\quad \left. + 955392 r^4 - 705536 r^3 - 387680 r^{11} + 1118472 r^8 - 1308960 r^7 + 175984 r^{12} \right. \\
&\quad \left. - 46176 r^{13} + 5120 r^{14} + 56016 \right] / \left[ 20 r^4 \right],
\end{aligned}$$

and the corresponding intervals are  $\mathcal{I}_1 = [1, 9/8)$ ,  $\mathcal{I}_2 = [9/8, 9/7)$ ,  $\mathcal{I}_3 = [9/7, 4/3)$ ,  $\mathcal{I}_4 = [4/3, 3/2)$ ,  $\mathcal{I}_5 = [3/2, 2)$ . See Figure 20 for the graph of  $\mu_S(r, \sqrt{3}/4)$  and  $\nu_S(r, \sqrt{3}/4)$ .

Then we get  $\text{HLAE}(r, \sqrt{3}/4) = \frac{(\mu_S(r, \sqrt{3}/4) - \mu(r))^2}{\nu_S(r, \sqrt{3}/4)}$  by substituting the relevant terms. See Figure 19.

With  $\epsilon = 2\sqrt{3}/7$ ,  $r \in [1, 3/2)$

$$\mu_S \left( r, 2\sqrt{3}/7 \right) = \begin{cases} -\frac{241}{54} r^2 + \frac{38}{3} r - 8 & \text{for } r \in [1, 9/7) \\ \frac{80 r^4 - 432 r^3 + 866 r^2 - 756 r + 243}{2 r^2} & \text{for } r \in [9/7, 3/2) \end{cases}$$

and

$$\nu_S \left( r, 2\sqrt{3}/7 \right) = \sum_{j=1}^6 \nu_j \left( r, 2\sqrt{3}/7 \right) \mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned}
\nu_1 \left( r, 2\sqrt{3}/7 \right) &= - \left[ 2495087 r^7 - 5067342 r^6 - 29145379 r^5 + 134149248 r^4 - 230713503 r^3 \right. \\
&\quad \left. + 202262778 r^2 - 90317349 r + 16336404 \right] / \left[ 14580 r \right], \\
\nu_2 \left( r, 2\sqrt{3}/7 \right) &= - \left[ 2495087 r^{10} - 5067342 r^9 - 29145379 r^8 + 134149248 r^7 - 140359071 r^6 \right. \\
&\quad \left. - 378587142 r^5 + 1465530651 r^4 - 2206303596 r^3 + 1786050000 r^2 - 765450000 r \right. \\
&\quad \left. + 136687500 \right] / \left[ 14580 r^4 \right], \\
\nu_3 \left( r, 2\sqrt{3}/7 \right) &= - \left[ 2495087 r^{10} - 5067342 r^9 - 29145379 r^8 + 134149248 r^7 - 309668679 r^6 \right. \\
&\quad \left. + 731864538 r^5 - 1559738349 r^4 + 2174176404 r^3 - 1767825000 r^2 + 765450000 r \right. \\
&\quad \left. - 136687500 \right] / \left[ 14580 r^4 \right], \\
\nu_4 \left( r, 2\sqrt{3}/7 \right) &= \left[ 1000147 r^8 - 654768 r^7 + 77561559 r^6 - 527363136 r^5 + 1468526760 r^4 \right. \\
&\quad \left. + 1767825000 r^2 - 765450000 r - 2157840000 r^3 + 136687500 + 24337 r^{10} \right. \\
&\quad \left. - 321426 r^9 \right] / \left[ 14580 r^4 \right], \\
\nu_5 \left( r, 2\sqrt{3}/7 \right) &= \frac{24337}{14580} r^6 - \frac{17857}{810} r^5 + \frac{1000147}{14580} r^4 - \frac{18188}{405} r^3 - \frac{174113}{1620} r^2 + \frac{8176}{45} r - 78, \\
\nu_6 \left( r, 2\sqrt{3}/7 \right) &= - \frac{(8 r^6 - 106 r^5 + 8709 r^4 - 39684 r^3 + 68000 r^2 - 51192 r + 14256) (2 r - 3)^4}{20 r^4}.
\end{aligned}$$

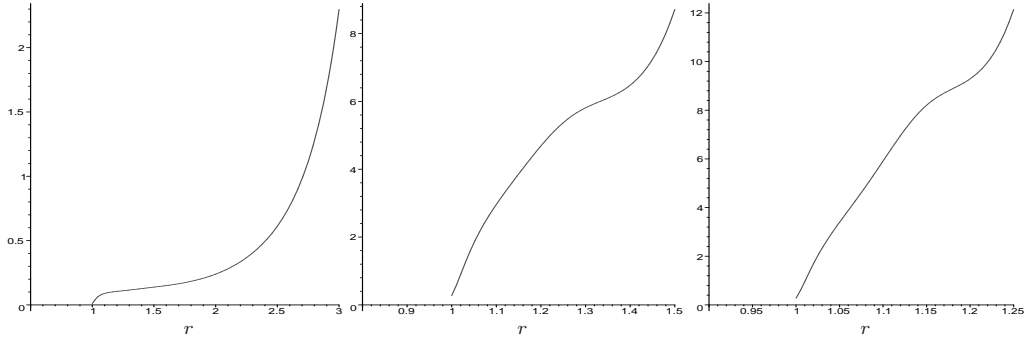


Figure 19: Hodges-Lehmann asymptotic efficacy against segregation alternative  $H_\epsilon^S$  as a function of  $r$  for  $\epsilon = \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$  (left to right).

The corresponding intervals are  $\mathcal{I}_1 = [1, 15/14)$ ,  $\mathcal{I}_2 = [15/14, 15/13)$ ,  $\mathcal{I}_3 = [15/13, 7/6)$ ,  $\mathcal{I}_4 = [7/6, 5/4)$ ,  $\mathcal{I}_5 = [5/4, 9/7)$ ,  $\mathcal{I}_6 = [9/7, 3/2)$ . See Figure 20 for the graph of  $\mu_S(r, 2\sqrt{3}/7)$  and  $\nu_S(r, 2\sqrt{3}/7)$ .

Then we get  $\text{HLAE}^S(r, 2\sqrt{3}/7) = \frac{(\mu_S(r, 2\sqrt{3}/7) - \mu_S(r, 2\sqrt{3}/7))^2}{\nu_S(r, 2\sqrt{3}/7)}$  by substituting the relevant terms. In Figure 19 are the graphs of  $\text{HLAE}^S(r, \epsilon)$  for  $\epsilon = \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$ .

From Figure 19, we see that, under  $H_\epsilon^S$ ,  $\text{HLAE}^S(r, \epsilon)$  appears to be an increasing function, dependent on  $\epsilon$ , of  $r$ . Let  $r_\delta(\epsilon)$  be the minimum  $r$  such that  $\rho_n(r)$  becomes degenerate under the alternative  $H_\epsilon^S$ . Then  $r_\delta(\sqrt{3}/8) = 4$ ,  $r_\delta(\sqrt{3}/4) = 2$ , and  $r_\delta(2\sqrt{3}/7) = 3/2$ . In fact, for  $\epsilon \in (0, \sqrt{3}/4]$ ,  $r_\delta(\epsilon) = \sqrt{3}/(2\epsilon)$  and for  $\epsilon \in (\sqrt{3}/4, \sqrt{3}/3)$ ,  $r_\delta(\epsilon) = \sqrt{3}/\epsilon - 2$ . Notice that  $\lim_{r \rightarrow r_\delta(\epsilon)} \text{HLAE}^S(\rho_n(r), \epsilon) = \infty$ , which is in agreement with PAE analysis because as  $\epsilon \rightarrow 0$  HLAE becomes PAE, and as  $\epsilon \rightarrow 0$ ,  $r_\delta(\epsilon) \rightarrow \infty$  and under  $H_0$ ,  $\rho_n(r)$  is degenerate for  $r = \infty$ . The above result for HLAE can also be generalized for arbitrary  $\epsilon$  as follows.

**Proposition 1** Let  $\tilde{r} := \text{argsup}_{r \in [1, r_\delta(\epsilon)]} \text{HLAE}^S(r, \epsilon)$  where  $r_\delta(\epsilon)$  is the value of  $r$  at which  $\rho_n(r)$  becomes degenerate under  $H_\epsilon^S$ . Then  $\tilde{r} = r_\delta(\epsilon)$ . In particular, for  $\epsilon \in [0, \sqrt{3}/4]$ ,  $r_\delta = \sqrt{3}/(2\epsilon)$  and for  $\epsilon \in (\sqrt{3}/4, \sqrt{3}/3]$ ,  $r_\delta = \sqrt{3}/\epsilon - 2$ .

**Proof:** Recall that  $\text{HLAE}^S(r, \epsilon) = \frac{(\mu_S(r, \epsilon) - \mu(r))^2}{\nu_S(r, \epsilon)}$ . For  $\epsilon \in [0, \sqrt{3}/4]$ ,  $\mu_S(r, \epsilon) \rightarrow 1$  and  $\nu(r, \epsilon) \rightarrow 0$  as  $r \rightarrow r_\delta(\epsilon) = \sqrt{3}/(2\epsilon)$ . Hence  $\text{HLAE}^S(r, \epsilon) \rightarrow \infty$  as  $r \rightarrow r_\delta(\epsilon) = \sqrt{3}/(2\epsilon)$ . So for  $\epsilon \in [0, \sqrt{3}/4]$ , the  $\tilde{r} = \sqrt{3}/(2\epsilon)$ . For  $\epsilon \in (\sqrt{3}/4, \sqrt{3}/3]$ , the result follows similarly. ■

So HLAE suggests choosing  $r$  larger as the segregation gets more severe, but choosing  $r$  too large will reduce power since  $r \geq r_\delta(\epsilon)$  guarantees the complete digraph under the alternative and, as  $r$  increases therefrom, provides an ever greater probability of seeing the complete digraph under the null.

In Figure 20, we plot the graphs of mean and asymptotic variance for  $r \in [1, 4]$  under segregation with  $\epsilon = 0, \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$ . Notice that  $\mu_S(r, \epsilon)$  gets larger as  $\epsilon$  gets larger at each  $r$  which is in agreement with the  $\mu_S(r, \epsilon)$  expressions in Section ???. However, the same ordering holds for  $\nu_S(r, \epsilon)$  at each  $r$  only for large  $r$ , but for small  $r$  the ordering is reversed. Furthermore, both the  $\sup_{r \in [1, \infty]} \nu_S(r, \epsilon)$  and  $\text{argsup}_{r \in [1, \infty]} \nu_S(r, \epsilon)$  seem to decrease as  $\epsilon$  increases.

### 3.8.2 Hodges-Lehmann Asymptotic Efficacy Under Association Alternatives

In the HLAE framework,  $\theta = \mathbf{E}_\epsilon^A[\rho_n(r)] = \mu_A(r, \epsilon)$  and  $\theta_0 = \mu_A(r, \epsilon = 0)$ . Then testing  $H_0 : \epsilon = 0$  versus  $H_a : \epsilon > 0$  is equivalent to  $H_0 : \mathbf{E}[\rho_n] = \mu(r)$  versus  $H_a : \mathbf{E}_\epsilon^A[\rho_n] = \mu_A(r, \epsilon) < \mu(r)$ . Let  $\delta = \frac{\mu(r, \epsilon) - \mu(r)}{\sigma_n(r, \epsilon)}$  and  $\tilde{R} = \frac{\rho_n(r) - \mu(r)}{\sigma_n(r, \epsilon)}$ , then  $\tilde{R} \xrightarrow{\mathcal{L}} \mathcal{N}(\delta, 1)$ .

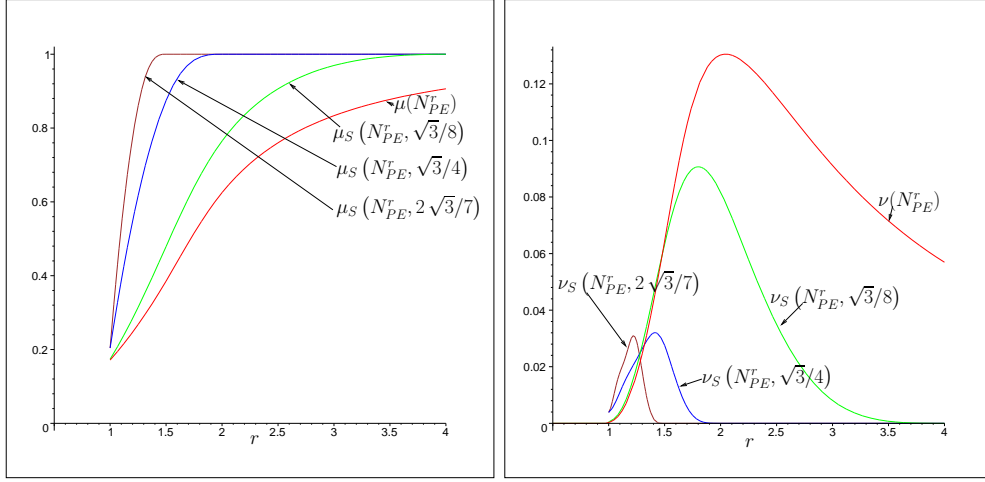


Figure 20: The mean  $\mu_S(r, \epsilon)$  (left) and asymptotic variance  $\nu_S(r, \epsilon)$  (right) as a function of  $r$  under segregation with  $\epsilon = 0, \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$ .

Hodges-Lehmann asymptotic efficacy (HLAE) (Hodges and Lehmann (1956)) is given by

$$\text{HLAE}^A(r, \epsilon) := \frac{(\mu_A(r, \epsilon) - \mu(r))^2}{\nu_A(r, \epsilon)}.$$

Rather than an arbitrary  $\epsilon$  we pick specific values:  $5\sqrt{3}/24, \sqrt{3}/12$ , and  $\sqrt{3}/21$ . Recall that  $\rho_n(r)$  is degenerate as  $r \rightarrow \infty$ . Furthermore  $\rho_n(r)$  is degenerate when  $r = 1$ .

With  $\epsilon = 5\sqrt{3}/24$ ,

$$\mu_A(r, 5\sqrt{3}/24) = \begin{cases} -\frac{1}{6}r^{-2} + \frac{1}{3} & \text{for } r \in [1, 3) \\ \frac{1}{3}r^2 - \frac{8}{3}r - \frac{55}{6}r^{-2} + \frac{19}{3} & \text{for } r \in [3, 4) \\ -\frac{55}{6}r^{-2} + 1 & \text{for } r \in [4, \infty) \end{cases}$$

and

$$\nu_A(r, 5\sqrt{3}/24) = \sum_{j=1}^5 \nu_j(r, 5\sqrt{3}/24) \mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned} \nu_1(r, 5\sqrt{3}/24) &= \frac{r^4 - 2r^2 + 1}{27r^6}, \\ \nu_2(r, 5\sqrt{3}/24) &= -\left[120r^{10} - 2176r^9 + 15340r^8 - 50304r^7 + 58754r^6 + 74880r^5 - 248577r^4 \right. \\ &\quad \left. + 138240r^3 + 47172r^2 + 23328r - 7305\right] / [405r^6], \\ \nu_3(r, 5\sqrt{3}/24) &= -\left[120r^{10} - 2176r^9 + 15180r^8 - 48960r^7 + 58754r^6 + 47440r^5 - 176547r^4 \right. \\ &\quad \left. + 138240r^3 - 70477r^2 + 23328r - 7305\right] / [405r^6], \\ \nu_4(r, 5\sqrt{3}/24) &= \left[10r^{12} - 192r^{11} + 1320r^{10} - 2944r^9 - 7590r^8 + 49920r^7 - 69986r^6 - 46480r^5 \right. \\ &\quad \left. + 184137r^4 - 143360r^3 + 71917r^2 - 23520r + 7315\right] / [405r^6], \\ \nu_5(r, 5\sqrt{3}/24) &= \frac{787r^4 - 7601r^2 - 16032r + 9265}{135r^6}. \end{aligned}$$

The corresponding intervals are  $\mathcal{I}_1 = (1, 3)$ ,  $\mathcal{I}_2 = [3, 7/2)$ ,  $\mathcal{I}_3 = [7/2, 2 + \sqrt{3})$ ,  $\mathcal{I}_4 = [2 + \sqrt{3}, 4)$ ,  $\mathcal{I}_5 = [4, \infty)$ . See Figure 22 for the graph of  $\mu_A(r, 5\sqrt{3}/24)$  and  $\nu_A(r, 5\sqrt{3}/24)$ .

Then we get  $\text{HLAE}^A(r, 5\sqrt{3}/24) = \frac{(\mu_A(r, 5\sqrt{3}/24) - \mu(r))^2}{\nu_A(r, 5\sqrt{3}/24)}$  by substituting the relevant terms. See Figure 21.

With  $\epsilon = \sqrt{3}/12$ ,

$$\mu_A(r, \sqrt{3}/12) = \begin{cases} \frac{6r^4 - 16r^3 + 18r^2 - 5}{18r^2} & \text{for } r \in [1, 2) \\ -\frac{37}{18}r^{-2} + 1 & \text{for } r \in [2, \infty) \end{cases}$$

and

$$\nu_A(r, \sqrt{3}/12) = \sum_{j=1}^3 \nu_j(r, \sqrt{3}/12) \mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned} \nu_1(r, \sqrt{3}/12) &= \left[ 10r^{12} - 96r^{11} + 240r^{10} + 192r^9 - 1830r^8 + 3360r^7 - 2650r^6 + 240r^5 + 1383r^4 \right. \\ &\quad \left. - 1280r^3 + 540r^2 - 144r + 35 \right] / [405r^6], \\ \nu_2(r, \sqrt{3}/12) &= \left[ 10r^{12} - 96r^{11} + 240r^{10} + 192r^9 - 1670r^8 + 2784r^7 - 2650r^6 + 2400r^5 - 1047r^4 \right. \\ &\quad \left. - 1280r^3 + 1269r^2 - 144r + 35 \right] / [405r^6], \\ \nu_3(r, \sqrt{3}/12) &= \frac{537r^4 - 683r^2 - 2448r + 1315}{405r^6}. \end{aligned}$$

The corresponding intervals are  $\mathcal{I}_1 = [1, 3/2)$ ,  $\mathcal{I}_2 = [3/2, 2)$ ,  $\mathcal{I}_3 = [2, \infty)$ . See Figure 22 for the graph of  $\mu(r, \sqrt{3}/12)$  and  $\nu(r, \sqrt{3}/12)$ .

Then we get  $\text{HLAE}^A(r, \sqrt{3}/12) = \frac{(\mu_A(r, \sqrt{3}/12) - \mu(r))^2}{\nu_A(r, \sqrt{3}/12)}$  by substituting the relevant terms. See Figure 21.

With  $\epsilon = \sqrt{3}/21$ ,

$$\mu_A(r, \sqrt{3}/21) = \begin{cases} \frac{7839r^4 - 27648r^3 + 49152r^2 - 35840r + 9216}{16200r^2} & \text{for } r \in [1, 8/7) \\ \frac{2719r^4 - 5592r^3 + 5760r^2 - 1536}{8100r^2} & \text{for } r \in (8/7, 3/2) \\ \frac{53r^4 + 2744r^3 - 7296r^2 + 8064r - 3104}{2700r^2} & \text{for } r \in (3/2, 12/7) \\ \frac{2719r^4 - 1440r^2 + 2112}{16200r^2} & \text{for } r \in (12/7, 7/4) \\ -\frac{2401r^4 - 73824r^2 + 153664r - 88548}{16200r^2} & \text{for } r \in (7/4, 2) \\ 1 - \frac{89}{54}r^{-2} & \text{for } r \in [2, \infty) \end{cases}$$

and

$$\nu_A(r, \sqrt{3}/21) = \sum_{j=1}^{10} \nu_j(r, \sqrt{3}/21) \mathbf{I}(\mathcal{I}_j)$$



where

$$\begin{aligned}
\nu_1 \left( r, \sqrt{3}/21 \right) &= \left[ 4124031 r^{12} - 22708224 r^{11} - 389826 r^{10} + 369129408 r^9 - 1592672721 r^8 \right. \\
&\quad + 3532359672 r^7 - 4721848374 r^6 + 4050858048 r^5 - 2387433568 r^4 + 995033088 r^3 \\
&\quad \left. - 209048784 r^2 - 43352064 r + 25952256 \right] / \left[ 65610000 r^6 \right], \\
\nu_2 \left( r, \sqrt{3}/21 \right) &= \left[ 6594660 r^{12} - 31178952 r^{11} - 14911074 r^{10} + 441735648 r^9 - 1578842961 r^8 \right. \\
&\quad + 3311083512 r^7 - 4669163574 r^6 + 4366966848 r^5 - 2522908768 r^4 + 778272768 r^3 \\
&\quad \left. - 93443280 r^2 + 14450688 r - 8650752 \right] / \left[ 65610000 r^6 \right], \\
\nu_3 \left( r, \sqrt{3}/21 \right) &= \left[ 826701 r^{12} - 7118748 r^{11} + 14155864 r^{10} + 18467640 r^9 - 104968680 r^8 \right. \\
&\quad + 165877272 r^7 - 128355690 r^6 + 27338184 r^5 + 47304144 r^4 - 52684800 r^3 \\
&\quad \left. + 24413592 r^2 - 7225344 r + 1966080 \right] / \left[ 32805000 r^6 \right], \\
\nu_4 \left( r, \sqrt{3}/21 \right) &= \left[ 826701 r^{12} - 7118748 r^{11} + 14155864 r^{10} + 18467640 r^9 + 20074008 r^8 \right. \\
&\quad - 671672808 r^7 + 2194076310 r^6 - 3382581816 r^5 + 2840904144 r^4 - 1262284800 r^3 \\
&\quad \left. + 240413592 r^2 - 7225344 r + 1966080 \right] / \left[ 32805000 r^6 \right], \\
\nu_5 \left( r, \sqrt{3}/21 \right) &= \left[ 826701 r^{12} - 7118748 r^{11} + 14155864 r^{10} + 18467640 r^9 - 137116617 r^8 \right. \\
&\quad + 512952192 r^7 - 1511673690 r^6 + 2773418184 r^5 - 2883095856 r^4 + 1560115200 r^3 \\
&\quad \left. - 335586408 r^2 - 7225344 r + 1966080 \right] / \left[ 32805000 r^6 \right], \\
\nu_6 \left( r, \sqrt{3}/21 \right) &= \left[ 826701 r^{12} - 7118748 r^{11} + 14155864 r^{10} + 18467640 r^9 - 91939401 r^8 \right. \\
&\quad + 125718912 r^7 - 128697690 r^6 + 139178184 r^5 - 60695856 r^4 - 52684800 r^3 \\
&\quad \left. + 48413592 r^2 - 7225344 r + 1966080 \right] / \left[ 32805000 r^6 \right], \\
\nu_7 \left( r, \sqrt{3}/21 \right) &= \left[ 226415 r^{12} - 1426740 r^{11} + 334536 r^{10} + 17196648 r^9 - 87678147 r^8 \right. \\
&\quad + 311364480 r^7 - 711864862 r^6 + 944809880 r^5 - 684036240 r^4 + 238099456 r^3 \\
&\quad \left. - 24048504 r^2 - 7633920 r + 4761088 \right] / \left[ 10935000 r^6 \right], \\
\nu_8 \left( r, \sqrt{3}/21 \right) &= \left[ 5786907 r^{12} - 42712488 r^{11} + 76274888 r^{10} + 51865788 r^8 - 300043296 r^7 \right. \\
&\quad + 132202536 r^6 + 171413760 r^5 - 93614976 r^4 + 147517440 r^3 - 194460480 r^2 \\
&\quad \left. + 67608576 r - 29061120 \right] / \left[ 262440000 r^6 \right], \\
\nu_9 \left( r, \sqrt{3}/21 \right) &= - \left[ 2470629 r^{12} - 25412184 r^{11} + 112001848 r^{10} - 1958438076 r^8 \right. \\
&\quad + 5449924256 r^7 + 6150612888 r^6 - 55820599296 r^5 + 109663683136 r^4 \\
&\quad \left. - 97335694848 r^3 + 40552466112 r^2 - 9825887232 r + 3078523200 \right] / \left[ 262440000 r^6 \right], \\
\nu_{10} \left( r, \sqrt{3}/21 \right) &= \frac{493829 r^4 - 433645 r^2 - 1765008 r + 929955}{455625 r^6}.
\end{aligned}$$

The corresponding intervals are  $\mathcal{I}_1 = (1, 2\sqrt{14}/7)$ ,  $\mathcal{I}_2 = [2\sqrt{14}/7, 8/7)$ ,  $\mathcal{I}_3 = [8/7, 5/4)$ ,  $\mathcal{I}_4 = [5/4, 4/3)$ ,  $\mathcal{I}_5 = [4/3, 10/7)$ ,  $\mathcal{I}_6 = [10/7, 3/2)$ ,  $\mathcal{I}_7 = [3/2, 12/7)$ ,  $\mathcal{I}_8 = [12/7, 7/4)$ ,  $\mathcal{I}_9 = [7/4, 2)$ ,  $\mathcal{I}_{10} = [2, \infty)$ . See Figure 22 for the graph of  $\mu_A(r, \sqrt{3}/21)$  and  $\nu_A(r, \sqrt{3}/21)$ .

Then we get  $\text{HLAE}^A(r, \sqrt{3}/21) = \frac{(\mu_A(r, \sqrt{3}/21) - \mu(r))^2}{\nu_A(r, \sqrt{3}/21)}$  by substituting the relevant terms. See Figure 21.

Notice that for  $\epsilon = 5\sqrt{3}/24$  and  $\epsilon = \sqrt{3}/12$ ,  $\arg\sup_{r \geq 1} \text{HLAE}^A(r, \epsilon) = 1$ . This result for HLAE can be generalized for arbitrary  $\epsilon$  as follows.

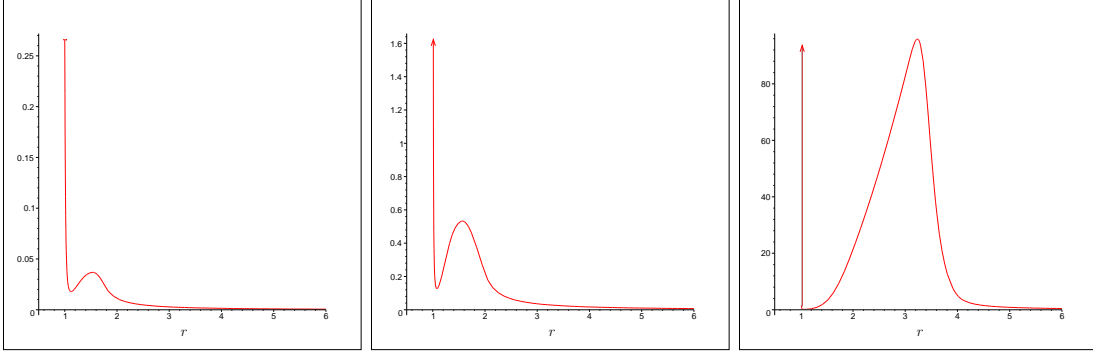


Figure 21: Hodges-Lehmann asymptotic efficacy against association alternative  $H_\epsilon^A$  as a function of  $r$  for  $\epsilon = \sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24$  (left to right).

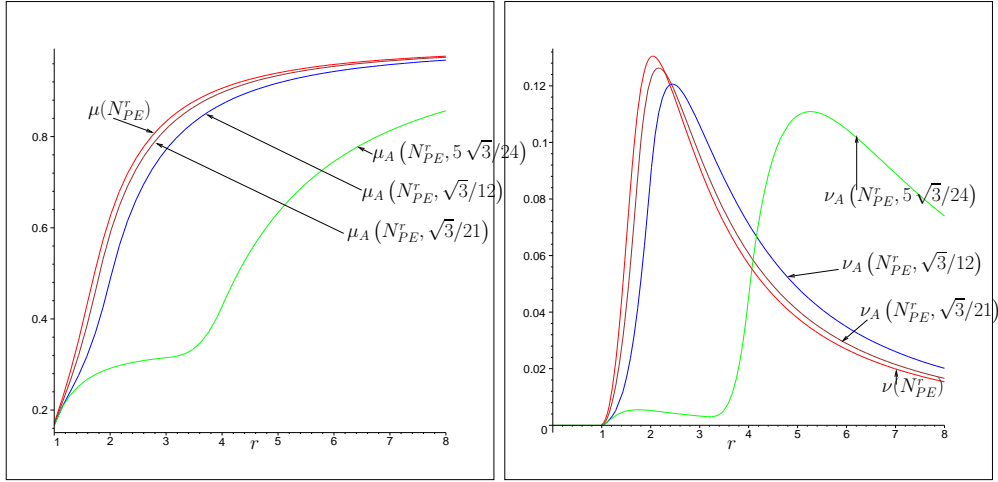


Figure 22: The mean  $\mu_A(r, \epsilon)$  (left) and asymptotic variance  $\nu_A(r, \epsilon)$  (right) as a function of  $r$  under association with  $\epsilon = 0, \sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24$ .

**Proposition 2** Let  $r^* := \operatorname{argsup}_{r \geq 1} \text{HLAE}^A(r, \epsilon)$  and  $\epsilon \geq \sqrt{3}/12$ . Then  $r^* = 1$ .

**Proof:** Recall that  $\text{HLAE}^A(r, \epsilon) = \frac{(\mu_A(r, \epsilon) - \mu(r))^2}{\nu_A(r, \epsilon)}$ . For  $\epsilon \in [\sqrt{3}/12, \sqrt{3}/3)$ ,  $\mu_A(r = 1, \epsilon) \rightarrow 1$  and  $\nu_A(r = 1, \epsilon) \rightarrow 0$ . Hence  $\text{HLAE}^A(r, \epsilon) \rightarrow \infty$  as  $r \rightarrow 1$ . So the desired result follows. ■

For  $\epsilon \in [0, \sqrt{3}/12]$ , it seems that for a while  $r^* = 1$  with respect to HLAE, e.g. for  $\epsilon = \sqrt{3}/21$ . But for sufficiently small  $\epsilon$ ,  $r^* > 1$  holds. This can also be seen as  $\epsilon \rightarrow 0$  in which case HLAE becomes PAE and the optimal value is about 1.006 with respect to PAE. Furthermore, observe that the  $\operatorname{argsup}$  for HLAE gets closer to 1 as  $\epsilon \rightarrow 0$  and  $\nu_A(r, \epsilon) > 0$  for  $\epsilon \in (0, \sqrt{3}/12)$  and  $\nu(r, \epsilon)$  gets larger as  $\epsilon \rightarrow 0$ .

Figure 21 contains a graph of HLAE against association as a function of  $r$  for  $\epsilon = 5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21$ . Notice that since  $\nu(r = 1, \epsilon) = 0$  for  $\epsilon \geq \sqrt{3}/12$ ,  $\text{HLAE}^A(r = 1, \epsilon) = \infty$  for  $\epsilon \geq \sqrt{3}/12$  and  $\lim_{r \rightarrow \infty} \text{HLAE}^A(r, \epsilon) = 0$ . In Figure 21 we see that, against  $H_\epsilon^A$ ,  $\text{HLAE}^A(r, \epsilon)$  has a local supremum for some  $r > 1$ . Let  $\tilde{r}_l$  be the value at which this local supremum is attained. Then  $\tilde{r}_l(5\sqrt{3}/24) \approx 3.2323$ ,  $\tilde{r}_l(\sqrt{3}/12) \approx 1.5676$ , and  $\tilde{r}_l(\sqrt{3}/21) \approx 1.533$ . Note that, as  $\epsilon$  gets smaller,  $\tilde{r}$  gets smaller. Furthermore,  $\text{HLAE}^A(r = 1, \sqrt{3}/21) < \infty$  and as  $\epsilon \rightarrow 0$ , so  $\tilde{r}$  becomes the global supremum, and  $\text{PAE}^A(r = 1) = 0$  and  $\operatorname{argsup}_{r \geq 1} \text{PAE}^A(r = 1) \approx 1.006$ . So HLAE suggests choosing moderate  $r$  when testing against association, whereas PAE suggests choosing small  $r$ .

Derivation of  $\mu_A(r, \epsilon)$  and  $\nu_A(r, \epsilon)$  for association with  $\epsilon = 5\sqrt{3}/24, \sqrt{3}/12$ , and  $\sqrt{3}/21$  are similar to — with the supports being the complements of — the corresponding segregation cases.

In Figure 22, we plot the graphs of mean and asymptotic variance for  $r \in [1, 8]$  under association with

$\epsilon = 0, \sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24$ . Notice that  $\mu_A(r, \epsilon)$  gets smaller as  $\epsilon$  gets larger at each  $r$  which is in agreement with the  $\mu_A(r, \epsilon)$  expressions in Appendix 2. However, the same ordering does not hold for  $\nu_A(r, \epsilon)$  at each  $r$ . For small  $r$  the ordering is same as in  $\mu_A(r, \epsilon)$ , but for large  $r$  the ordering is reversed. Furthermore,  $\sup_{r \in [1, \infty]} \nu_A(r, \epsilon)$  seems to decrease as  $\epsilon$  increases while  $\arg\sup_{r \in [1, \infty]} \nu_A(r, \epsilon)$  seems to increase as  $\epsilon$  increases.

### 3.9 Asymptotic Power Function Analysis

The asymptotic power function (see, e.g., Kendall and Stuart (1979)) can also be investigated as a function of  $r$ ,  $n$ , and  $\epsilon$  using the asymptotic critical value and an appeal to normality.

#### 3.9.1 Asymptotic Power Function Analysis Under Segregation

Under segregation, for sufficiently large  $n$ , we reject  $H_0$  when  $\sqrt{n} \left( \frac{\rho_n(r) - \mu_S(r, \epsilon)}{\sqrt{\nu_S(r, \epsilon)}} \right) > z_{(1-\alpha)}$  where  $z_{(1-\alpha)}$  is the  $(1 - \alpha) \times 100$  percentile of the standard normal distribution, e.g. with  $\alpha = .05$ ,  $z_{.95} \approx 1.645$ . Then size  $\alpha$  critical region for large samples is

$$\rho_n(r) > \mu(r) + z_{(1-\alpha)} \cdot \sqrt{\nu(r)/n}.$$

Under a specific segregation alternative  $H_\epsilon^S$ , the asymptotic power function is given by

$$\begin{aligned} \Pi_S(r, n, \epsilon) &:= P\left(\rho_n(r) > \mu(r) + z_{(1-\alpha)} \cdot \sqrt{\nu(r)/n}\right) \\ &= 1 - \Phi\left(\frac{z_{(1-\alpha)} \sqrt{\nu(r)}}{\sqrt{\nu_S(r, \epsilon)}} + \frac{\sqrt{n}(\mu(r) - \mu_S(r, \epsilon))}{\sqrt{\nu_S(r, \epsilon)}}\right). \end{aligned}$$

With  $\epsilon = \sqrt{3}/8$ ,  $\Pi_S(r, n, \epsilon)$  at level  $\alpha = .05$  is plotted in Figure 23. Observe that  $\Pi_S(r, n, \sqrt{3}/8) \rightarrow 0$  as  $r \rightarrow 4$  for  $n = 5, 10, 15$ . Let  $r_g^*(n, \epsilon)$  be the value at which  $\Pi_S(r, n, \epsilon)$  attains its global supremum and  $r_l^*(n, \epsilon)$  be the value at which  $\Pi_S(r, n, \epsilon)$  attains its local supremum. Then  $r_g^*(5, \sqrt{3}/8) \approx 1.260$ ,  $r_g^*(10, \sqrt{3}/8) \approx 1.3741$  and  $r_l^*(10, \sqrt{3}/8) \approx 2.3818$ ,  $r_g^*(15, \sqrt{3}/8) \approx 3.3724$  and  $r_l^*(15, \sqrt{3}/8) \approx 1.45$ ,  $r_g^*(20, \sqrt{3}/8) = 4$  and  $r_l^*(20, \sqrt{3}/8) \approx 1.5$ . Finally,  $r_g^*(n, \sqrt{3}/8) = 4$  for  $n = 20, 50, 100$  and  $\Pi_S(r, n, \sqrt{3}/8)$  has a hump for  $n = 10$  and  $n = 15$ .

With  $\epsilon = \sqrt{3}/4$ ,  $\Pi_S(r, n, \epsilon)$  at level  $\alpha = .05$  is plotted in Figure 24. Observe that  $\Pi_S(r, n, \sqrt{3}/4) \rightarrow 1$  as  $r \rightarrow 2$  for  $n = 3, 5$ . Moreover,  $r_g^*(n, \sqrt{3}/4) = 2$ , for  $n = 3, 5$ .

With  $\epsilon = 2\sqrt{3}/7$ ,  $\Pi_S(r, n, \epsilon)$  at level  $\alpha = .05$  is plotted in Figure 25. Observe that  $\Pi_S(r, n, 2\sqrt{3}/7) \rightarrow 1$  as  $r \rightarrow 3/2$  for  $n = 3, 5$  and  $r_g^*(n, 2\sqrt{3}/7) = 2$ , for  $n = 3, 5$ .

#### 3.9.2 Asymptotic Power Function Analysis Under Association

Under association, for sufficiently large  $n$ , we reject  $H_0$  when  $\sqrt{n} \left( \frac{\rho_n(r) - \mu_A(r, \epsilon)}{\sqrt{\nu_A(r, \epsilon)}} \right) < z_\alpha$  where  $z_\alpha$  is the  $\alpha \times 100$  percentile of the standard normal distribution, e.g. with  $\alpha = .05$ ,  $z_{.05} \approx -1.645$ . Then size  $\alpha$  critical region for large samples is

$$\rho_n(r) < \mu(r) + z_\alpha \cdot \sqrt{\nu(r)/n}.$$

Under  $H_\epsilon^A$ , we have

$$\begin{aligned} \Pi_A(r, n, \epsilon) &:= P\left(\rho_n(r) < \mu(r) + z_\alpha \cdot \sqrt{\nu(r)/n}\right) \\ &= \Phi\left(\frac{z_\alpha \sqrt{\nu(r)}}{\sqrt{\nu_A(r, \epsilon)}} + \frac{\sqrt{n}(\mu(r) - \mu_A(r, \epsilon))}{\sqrt{\nu_A(r, \epsilon)}}\right). \end{aligned}$$

With  $\epsilon = \sqrt{3}/21$ ,  $\Pi_A(r, n, \epsilon)$  at level  $\alpha = .05$  is plotted in Figure 26. Observe that  $\Pi_A(r, n, \sqrt{3}/21) \rightarrow .057$  as  $r \rightarrow \infty$  for  $n = 5, 10, 100$ . Let  $\hat{r}(n, \epsilon)$  be the value at which  $\Pi_A(r, n, \epsilon)$  attains its supremum. Then,

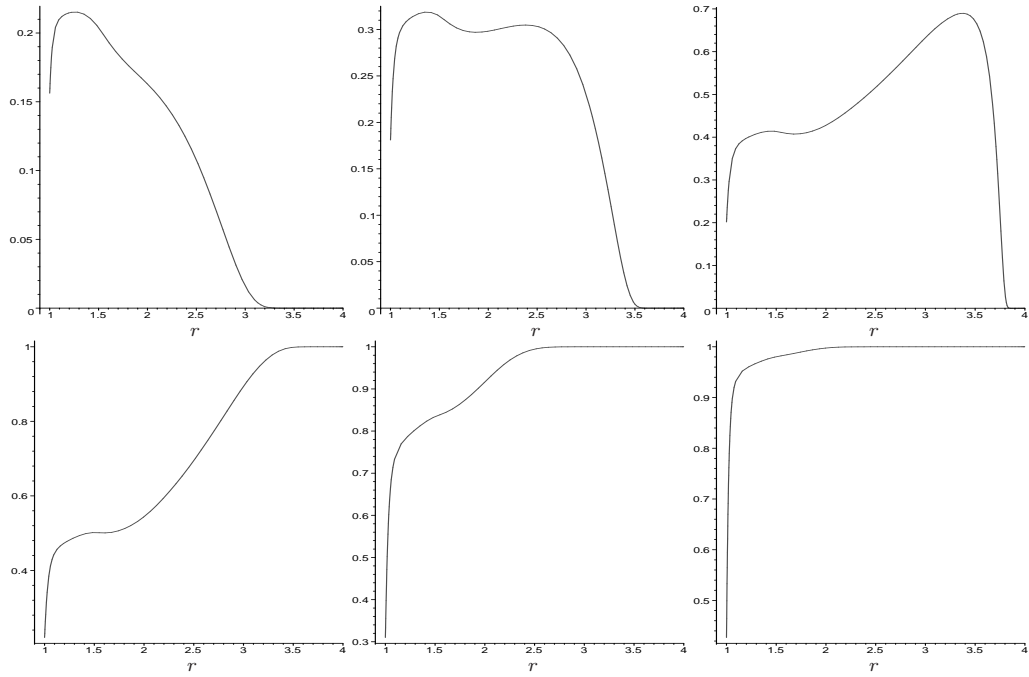


Figure 23: Asymptotic power function against segregation alternative  $H_{\sqrt{3}/8}^S$  as a function of  $r$  for  $n = 5, 10, 15, 20, 50, 100$ .

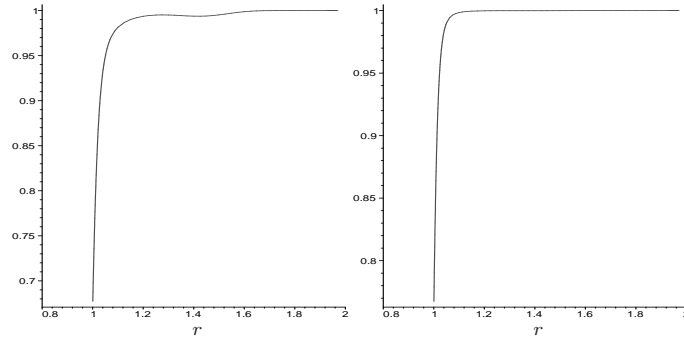


Figure 24: Asymptotic power function against segregation alternative  $H_{\sqrt{3}/4}^S$  as a function of  $r$  for  $n = 3$  (left) and  $n = 5$  (right).

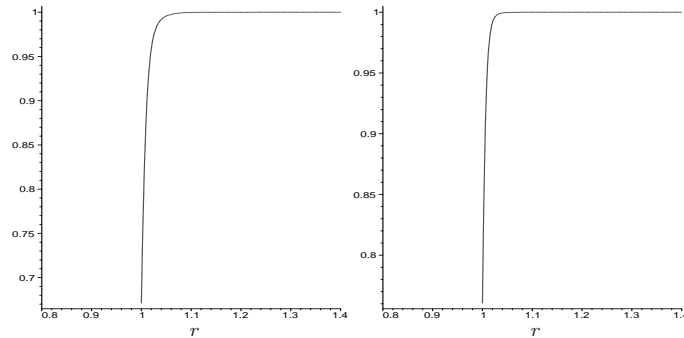


Figure 25: Asymptotic power function against segregation alternative  $H_{\sqrt{3}/7}^S$  as a function of  $r$  for  $n = 3$  (left) and  $n = 5$  (right).

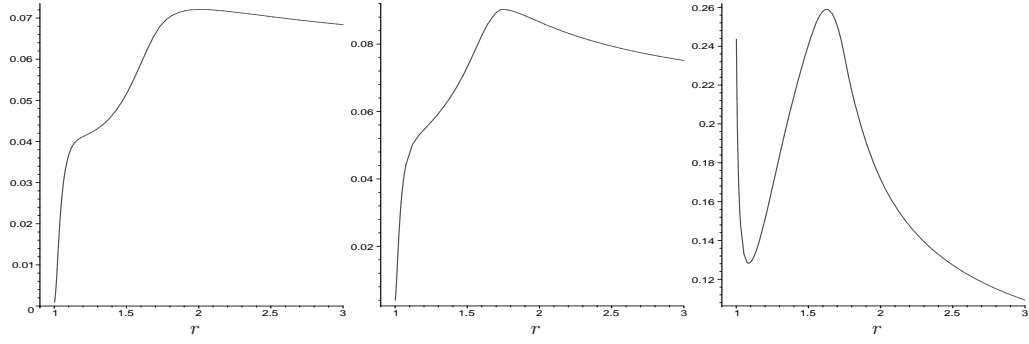


Figure 26: Asymptotic power function against association alternative  $H^A_{\sqrt{3}/21}$  as a function of  $r$  for  $n = 5, 10, 100$

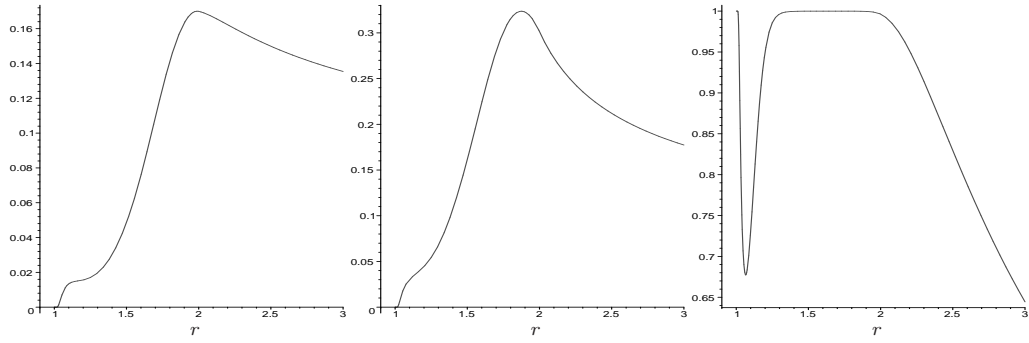


Figure 27: Asymptotic power function against association alternative  $H^A_{\sqrt{3}/12}$  as a function of  $r$  for  $n = 5, 10, 100$

$\hat{r}(5, \sqrt{3}/21) \approx 2.01$ , and  $\hat{r}(10, \sqrt{3}/21) \approx 1.875$ , and  $\hat{r}(100, \sqrt{3}/21) \approx 1.645$ . Moreover,  $\Pi_A(r, 100, \sqrt{3}/21)$  attains a local infimum at  $\approx 1.065$ .

With  $\epsilon = \sqrt{3}/12$ ,  $\Pi_A(r, n, \epsilon)$  at level  $\alpha = .05$  is plotted in Figure 27. Observe that  $\Pi_A(r, n, \sqrt{3}/12) \rightarrow .0766$  as  $r \rightarrow \infty$  for  $n = 5, 10, 100$ . Moreover,  $\hat{r}(5, \sqrt{3}/12) \approx 1.99$ ,  $\hat{r}(10, \sqrt{3}/12) \approx 1.75$ , and  $\hat{r}(100, \sqrt{3}/12) \approx 1.60$ . Moreover,  $\Pi_A(r, 100, \sqrt{3}/21)$  attains a local infimum at  $\approx 1.105$ .

With  $\epsilon = 5\sqrt{3}/24$ ,  $\Pi_A(r, n, \epsilon)$  at level  $\alpha = .05$  is plotted in Figure 28. Observe that  $\Pi(r, n, 5\sqrt{3}/24) \rightarrow 1$  as  $r \rightarrow 2$  for  $n = 5, 10$ .

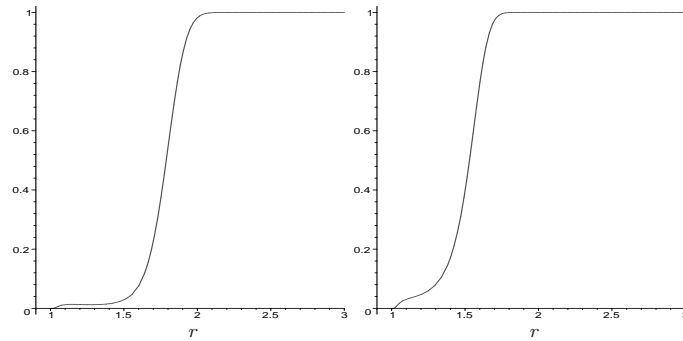


Figure 28: Asymptotic power function against association alternative  $H^A_{5\sqrt{3}/24}$  as a function of  $r$  for  $n = 5, 10$ .

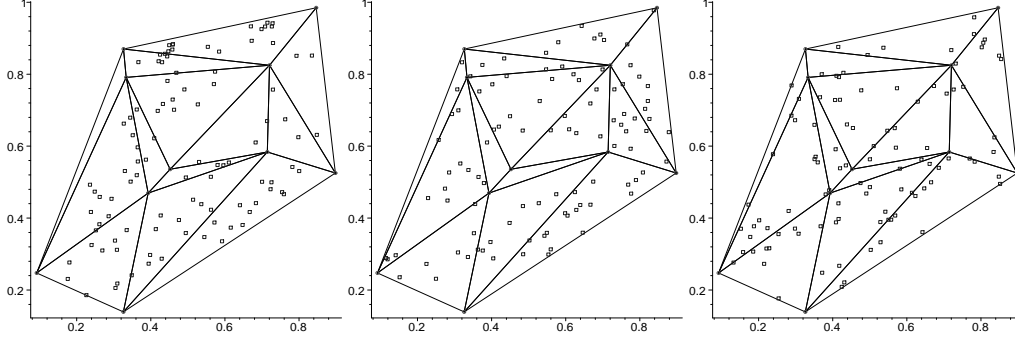


Figure 29: Realization of segregation (left),  $H_0$  (middle), and association (right) for  $|\mathcal{Y}| = 10$ ,  $J = 13$ , and  $n = 100$ .

### 3.10 Multiple Triangle Case

Suppose  $\mathcal{Y}$  is a finite collection of points in  $\mathbb{R}^2$  with  $|\mathcal{Y}| \geq 3$ . Consider the Delaunay triangulation (assumed to exist) of  $\mathcal{Y}$ , where  $T_j$  denotes the  $j^{\text{th}}$  Delaunay triangle,  $J$  denotes the number of triangles, and  $C_H(\mathcal{Y})$  denotes the convex hull of  $\mathcal{Y}$ . We wish to test

$$H_0 : X_i \stackrel{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}))$$

against segregation and association alternatives.

Figure 29 and Figure 30 are graphs of realizations of  $n = 100$  and  $n = 1000$  observations which are independent and identically distributed according to  $\mathcal{U}(C_H(\mathcal{Y}))$  for  $|\mathcal{Y}| = 10$  and  $J = 13$  and realizations of  $n = 100$  and  $n = 1000$  observations under segregation and association for the same  $\mathcal{Y}$ , respectively.

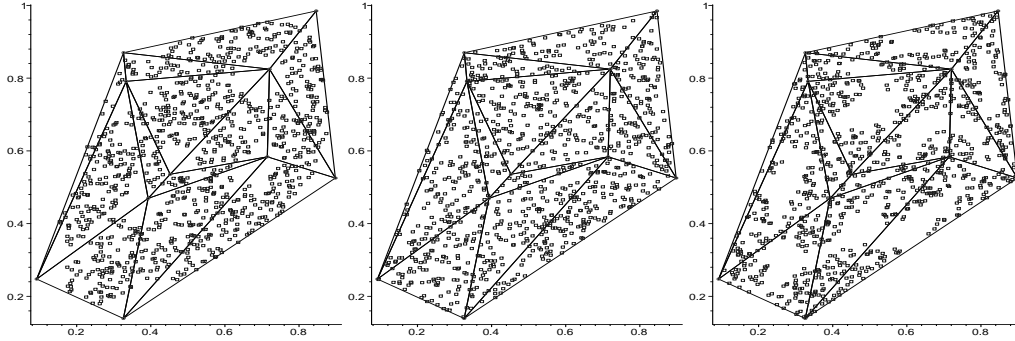


Figure 30: Realization of segregation (left),  $H_0$  (middle), and association (right) for  $|\mathcal{Y}| = 10$ ,  $J = 13$ , and  $n = 1000$ .

The digraph  $D$  is constructed using  $N_{\mathcal{Y}_j}^r(\cdot)$  as described in Section 2.3, here for  $X_i \in T_j$  the three points in  $\mathcal{Y}$  defining the Delaunay triangle  $T_j$  are used as  $\mathcal{Y}_j$ . Let  $\rho_n(r, J)$  be the relative density of the digraph based on  $\mathcal{X}_n$  and  $\mathcal{Y}$  which yields  $J$  Delaunay triangles, and let  $w_j = A(T_j)/A(C_H(\mathcal{Y}))$  for  $j = 1, \dots, J$ , where  $A(C_H(\mathcal{Y})) = \sum_{j=1}^J A(T_j)$  with  $A(\cdot)$  being the area functional. Then we obtain the following as a corollary to Theorem 2.

**Corollary 1** The asymptotic null distribution for  $\rho_n(r, J)$  conditional on  $\mathcal{W} = \{w_1, \dots, w_J\}$  for  $r \in [1, \infty)$  is given by  $\mathcal{N}(\mu(r, J), \nu(r, J)/n)$  provided that  $\nu(r, J) > 0$  with

$$\mu(r, J) := \mu(r) \sum_{j=1}^J w_j^2 \quad \text{and} \quad \nu(r, J) := \nu(r) \sum_{j=1}^J w_j^3 + 4\mu(r)^2 \left[ \sum_{j=1}^J w_j^3 - \left( \sum_{j=1}^J w_j^2 \right)^2 \right], \quad (14)$$

where  $\mu(r)$  and  $\nu(r)$  are given in Equations (8) and (9), respectively.

**Proof:** See Appendix 5 for the proof ■

$r$	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5	10
$n = 100, N = 1000$										
$\hat{\alpha}_S(n, J)$	.144	.141	.124	.101	.095	.087	.070	.075	.071	.072
$\hat{\beta}_n^S(r, \sqrt{3}/8, J)$	.191	.383	.543	.668	.714	.742	.742	.625	.271	.124
$\hat{\alpha}_A(n, J)$	.118	.111	.089	.081	.065	.062	.067	.064	.068	.071
$\hat{\beta}_n^A(r, \sqrt{3}/12, J)$	.231	.295	.356	.338	.269	.209	.148	.095	.113	.167
$n = 200, N = 1000$										
$\hat{\alpha}_S(n, J)$	.095	.092	.087	.077	.073	.076	.072	.071	.074	.073
$\hat{\beta}_n^S(r, \sqrt{3}/8, J)$	.135	.479	.743	.886	.927	.944	.959	.884	.335	.105
$\hat{\alpha}_A(n, J)$	.071	.071	.062	.057	.055	.047	.038	.035	.036	.040
$\hat{\beta}_n^A(r, \sqrt{3}/12, J)$	.182	.317	.610	.886	.952	.985	.972	.386	.143	.068
$n = 500, N = 1000$										
$\hat{\alpha}_S(n, J)$	.089	.092	.087	.086	.080	.078	.079	.079	.076	.081
$\hat{\beta}_n^S(r, \sqrt{3}/8, J)$	.145	.810	.981	.997	.999	1.000	1.000	1.000	.604	.130
$\hat{\alpha}_A(n, J)$	.087	.085	.076	.075	.073	.075	.072	.067	.066	.061
$\hat{\beta}_n^A(r, \sqrt{3}/12, J)$	.241	.522	.937	1.000	1.000	1.000	1.000	.712	.187	.063

Table 7: The empirical significance level and empirical power values under  $H_{\sqrt{3}/8}^S$  and  $H_{\sqrt{3}/12}^A$ ,  $N = 1000$ ,  $n = 100$ , and  $J = 13$ , at  $\alpha = .05$  for the realization of  $\mathcal{Y}$  in Figure 30.

By an appropriate application of Jensen's inequality, we see that  $\sum_{j=1}^J w_j^3 \geq \left(\sum_{j=1}^J w_j^2\right)^2$ . Therefore, the covariance  $\nu(r, J) = 0$  iff both  $\nu(r) = 0$  and  $\sum_{j=1}^J w_j^3 = \left(\sum_{j=1}^J w_j^2\right)^2$  hold, so asymptotic normality may hold even when  $\nu(r) = 0$ .

Similarly, for the segregation (association) alternatives with  $4 \epsilon^2/3 \times 100\%$  of the triangles around the vertices of each triangle is forbidden (allowed), we obtain the above asymptotic distribution of  $\rho_n(r, J)$  with  $\mu(r, J)$  being replaced by  $\mu(r, J, \epsilon)$ ,  $\nu(r, J)$  by  $\nu(r, J, \epsilon)$ ,  $\mu(r)$  by  $\mu(r, \epsilon)$ , and  $\nu(r)$  by  $\nu(r, \epsilon)$ .

Thus in the case of  $J > 1$ , we have a (conditional) test of  $H_0 : X_i \stackrel{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}))$  which once again rejects against segregation for large values of  $\rho_n(r, J)$  and rejects against association for small values of  $\rho_n(r, J)$ .

The segregation (with  $\delta = 1/16$  i.e.  $\epsilon = \sqrt{3}/8$ ), null, and association (with  $\delta = 1/4$  i.e.  $\epsilon = \sqrt{3}/12$ ) realizations (from left to right) are depicted in Figure 29 with  $n = 100$  and in Figure 30 with  $n = 1000$ . For both  $n = 100$  and  $n = 1000$ , for the null realization, the  $p$ -value is greater than 0.1 for all  $r$  values and both alternatives. For the segregation realization, with  $n = 100$  we obtain  $p < 0.001$  for  $r \leq 3$ ,  $p = 0.025$  for  $r = 5$ , and  $p > 0.1$  for  $r \geq 10$  and with  $n = 1000$  we obtain  $p < 0.0031$  for  $1 < r \leq 5$  and  $p > 0.24$  for  $r = 1$  and  $r \geq 10$ . For the association realization, with  $n = 100$ , we obtain  $p < 0.05$  for  $r = 1.5, 2$ , and  $p > 0.06$  for other values of  $r$  and with  $n = 1000$ , we obtain  $p < 0.0135$  for  $1 < r \leq 3$ ,  $p = .14$  for  $r = 1$ , and  $p > 0.25$  for  $r \geq 5$ . Note that this is only for one realization of  $\mathcal{X}_n$ .

We implement the above described Monte Carlo experiment 1000 times with  $n = 100$ ,  $n = 200$ , and  $n = 500$  and find the empirical significance levels  $\hat{\alpha}_S(n, J)$  and  $\hat{\alpha}_A(n, J)$  and the empirical powers  $\hat{\beta}_n^S(r, \sqrt{3}/8, J)$  and  $\hat{\beta}_n^A(r, \sqrt{3}/12, J)$ . These empirical estimates are presented in Table 7 and plotted in Figures 31 and 32. Notice that the empirical significance levels are all larger than .05 for both alternatives, so this test is liberal in rejecting  $H_0$  against both alternatives for the given realization of  $\mathcal{Y}$  and  $n$  values. The smallest empirical significance levels and highest empirical power estimates occur at moderate  $r$  values ( $r = 3/2, 2, 3$ ) against segregation and at smaller  $r$  values ( $r = \sqrt{2}, 3/2$ ) against association. Based on this analysis, for the given realization of  $\mathcal{Y}$ , we suggest the use of moderate  $r$  values for segregation and slightly smaller for association. Notice also that as  $n$  increases, the empirical power estimates gets larger for both alternatives.

**Remark** The conditional test presented here is appropriate when  $w_j \in \mathcal{W}$  are fixed, not random. An unconditional version requires the joint distribution of the number and relative size of Delaunay triangles when  $\mathcal{Y}$  is, for instance, a Poisson point pattern. Alas, this joint distribution is not available (see Okabe et al. (2000)).



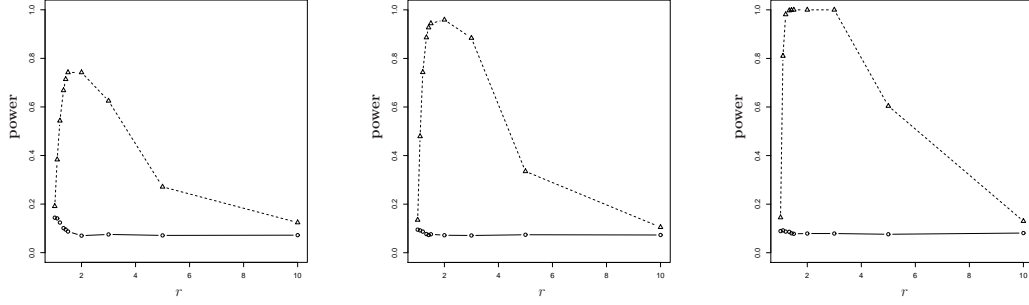


Figure 31: Monte Carlo power using the asymptotic critical value against  $H^S_{\sqrt{3}/8}$ , as a function of  $r$ , for  $n = 100$  (left),  $n = 200$  (middle), and  $n = 500$  (right) conditional on the realization of  $\mathcal{Y}$  in Figure 29. The circles represent the empirical significance levels while triangles represent the empirical power values.

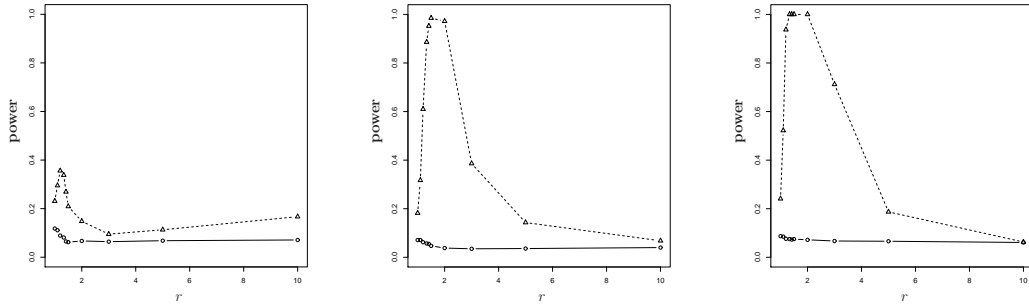


Figure 32: Monte Carlo power using the asymptotic critical value against  $H^A_{\sqrt{3}/12}$  as a function of  $r$ , for  $n = 100$  (left),  $n = 200$  (middle), and  $n = 500$  (right) conditional on the realization of  $\mathcal{Y}$  in Figure 30. The circles represent the empirical significance levels while triangles represent the empirical power values.

### 3.10.1 Related Test Statistics in Multiple Triangle Case

For  $J > 1$ , we have derived the asymptotic distribution of  $\rho_n(r, J) = \frac{|\mathcal{A}|}{n(n-1)}$ . Let  $\mathcal{A}_j$  be the number of arcs and  $\rho_{n_j}(r)$  be the relative density for triangle  $T_j$  and  $n_j := |\mathcal{X}_n \cap T_j|$ , for  $j = 1, \dots, J$ . So

$$\sum_{j=1}^J \frac{n_j(n_j-1)}{n(n-1)} \rho_{n_j}(r) = \rho_n(r, J),$$

since

$$\sum_{j=1}^J \frac{n_j(n_j-1)}{n(n-1)} \rho_{n_j}(r) = \frac{\sum_{j=1}^J |\mathcal{A}_j|}{n(n-1)} = \frac{|\mathcal{A}|}{n(n-1)} = \rho_n(r, J).$$

Let  $\hat{U}_n := \sum_{j=1}^J w_j^2 \cdot \rho_{n_j}(r)$  where  $w_j = A(T_j)/A(C_H(\mathcal{Y}))$ . Since  $\rho_{n_j}(r)$  are asymptotically independent,  $\sqrt{n}(\hat{U}_n - \mu(r, J))$  and  $\sqrt{n}(\rho_n(r, J) - \mu(r, J))$  both converge in distribution to  $\mathcal{N}(0, \nu(r, J))$ .

In the denominator of  $\rho_n(r, J)$ , we use  $n(n-1)$  as the maximum number of arcs possible. However, by definition, we can at most have a digraph with  $J$  complete symmetric components of order  $n_j$ , for  $j = 1, \dots, J$ . Then the maximum number possible is  $n_t := \sum_{j=1}^J n_j(n_j-1)$ . So the (adjusted) relative density is  $\rho_{n,J}^{adj}(r) := \frac{|\mathcal{A}|}{n_t}$  and  $\rho_n^{adj}(r) = \frac{\sum_{j=1}^J |\mathcal{A}_j|}{n_t} = \sum_{j=1}^J \frac{n_j(n_j-1)}{n_t} \rho_{n_j}(r)$ . Since  $\frac{n_j(n_j-1)}{n_t} \geq 0$  for each  $j$ , and  $\sum_{j=1}^J \frac{n_j(n_j-1)}{n_t} = 1$ ,  $\rho_{n,J}^{adj}(r)$  is a mixture of  $\rho_{n_j}(r)$ . Then  $\mathbf{E}[\rho_{n,J}^{adj}(r)] = \mu(r, J)$  and the asymptotic variance of  $\rho_{n,J}^{adj}(r)$  is

$$\frac{1}{n} \left[ \nu(r) \left( \sum_{j=1}^J w_j^3 / \left( \sum_{j=1}^J w_j^2 \right)^2 \right) + 4\mu(r)^2 \left( \sum_{j=1}^J w_j^3 / \left( \sum_{j=1}^J w_j^2 \right)^2 - 1 \right) \right]. \square$$

### 3.10.2 Asymptotic Efficacy Analysis for $J > 1$

The PAE, HLAE, and asymptotic power function analysis are given for  $J = 1$ . For  $J > 1$ , the analysis will depend on both the number of triangles as well as the relative sizes of the triangles. So the optimal  $r$  values with respect to these efficacy criteria for  $J = 1$  do not necessarily hold for  $J > 1$ , so the analysis need to be updated, given the values of  $J$  and  $\mathcal{W}$ .

Under segregation alternative  $H_\epsilon^S$ , the PAE is given by

$$\text{PAE}_J^S(r) = \frac{(\mu_S''(r, J, \epsilon = 0))^2}{\nu(r, J)} = \frac{(\mu_S''(r, \epsilon = 0) \sum_{j=1}^J w_j^2)^2}{\nu(r) \sum_{j=1}^J w_j^3 + 4\mu(r)^2 \left( \sum_{j=1}^J w_j^3 - \left( \sum_{j=1}^J w_j^2 \right)^2 \right)}. \quad (15)$$

Under association alternative  $H_\epsilon^A$  the PAE is similar.

In Figure 33, we present the PAE as a function of  $r$  for both segregation and association conditional on the realization of  $\mathcal{Y}$  in Figure 30. Notice that, unlike  $J = 1$  case,  $\text{PAE}_J^S(r)$  is bounded. Some values of interest are  $\text{PAE}_J^S(\rho_n(1)) = .3884$ ,  $\lim_{r \rightarrow \infty} \text{PAE}_J^S(r) = \frac{8 \sum_{j=1}^J w_j^2}{256 \left( \sum_{j=1}^J w_j^3 - \left( \sum_{j=1}^J w_j^2 \right)^2 \right)} \approx 139.34$ ,  $\text{argsup}_{r \in [1, 2]} \text{PAE}_J^S(r) \approx 1.974$ . As for association,  $\text{PAE}_J^A(r = 1) = 422.9551$ ,  $\lim_{r \rightarrow \infty} \text{PAE}_J^A(r) = 0$ ,  $\text{argsup}_{r \geq 1} \text{PAE}_J^A(r) \approx 1.5$  with  $\text{PAE}_J^A(r = 1.5) \approx 1855.9672$ . Based on the asymptotic efficacy analysis, we suggest, for large  $n$  and small  $\epsilon$ , choosing moderate  $r$  for testing against segregation and association.

Under segregation, the HLAE is given by

$$\text{HLAE}_J^S(r, \epsilon) := \frac{(\mu_S(r, J, \epsilon) - \mu(r, J))^2}{\nu_S(r, J, \epsilon)} = \frac{(\mu_S(r, \epsilon) \left( \sum_{j=1}^J w_j^2 \right) - \mu(r) \left( \sum_{j=1}^J w_j^2 \right))^2}{\nu_S(r, \epsilon) \sum_{j=1}^J w_j^3 + 4\mu_S(r, \epsilon)^2 \left( \sum_{j=1}^J w_j^3 - \left( \sum_{j=1}^J w_j^2 \right)^2 \right)}. \quad (16)$$

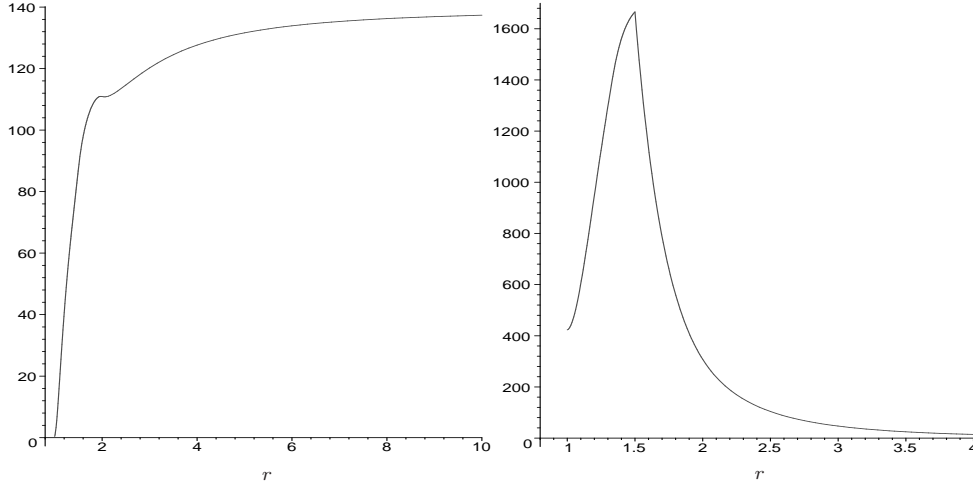


Figure 33: Pitman asymptotic efficacy against segregation (left) and association (right) as a function of  $r$  with the realization of  $\mathcal{Y}$  in Figure 29. Notice that vertical axes are differently scaled.

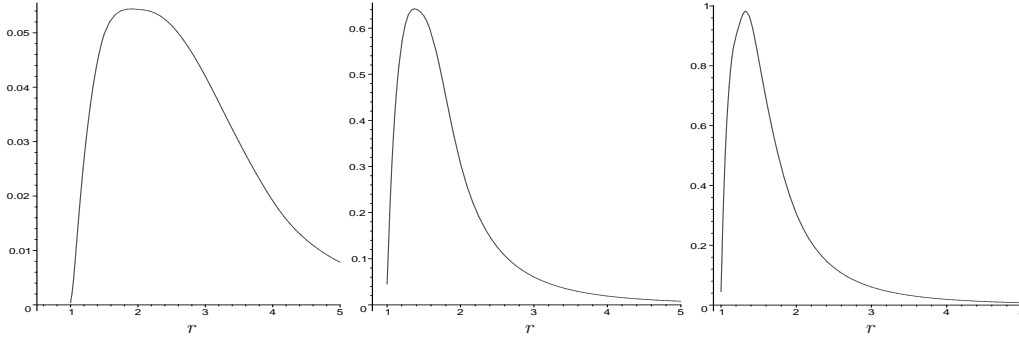


Figure 34: Hodges-Lehmann asymptotic efficacy against segregation alternative  $H_\epsilon^S$  as a function of  $r$  for  $\epsilon = \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7$  (left to right) conditional on the realization of  $\mathcal{Y}$  in Figure 29.

Notice that  $\text{HLAE}_J^S(r, \epsilon = 0) = 0$  and  $\lim_{r \rightarrow \infty} \text{HLAE}_J^S(r, \epsilon) = 0$ .

We calculate HLAE of  $\rho_n(r, J)$  under  $H_\epsilon^S$  for  $\epsilon = \sqrt{3}/8$ ,  $\epsilon = \sqrt{3}/4$ , and  $\epsilon = 2\sqrt{3}/7$ . In Figure 34 we present  $\text{HLAE}_J^S(r, \epsilon)$  for these  $\epsilon$  values conditional on the realization of  $\mathcal{Y}$  in Figure 29.

Note that with  $\epsilon = \sqrt{3}/8$ ,  $\text{HLAE}_J^S(r = 1, \sqrt{3}/8) \approx .0004$  and  $\text{argsup}_{r \in [1, \infty]} \text{HLAE}_J^S(r, \sqrt{3}/8) \approx 1.8928$  with the supremum  $\approx .0544$ . With  $\epsilon = \sqrt{3}/4$ ,  $\text{HLAE}_J^S(r = 1, \sqrt{3}/4) \approx .0450$  and  $\text{argsup}_{r \in [1, \infty]} \text{HLAE}_J^S(r, \sqrt{3}/4) \approx 1.3746$  with the supremum  $\approx .6416$ . With  $\epsilon = 2\sqrt{3}/7$ ,  $\text{HLAE}_J^S(r = 1, 2\sqrt{3}/7) \approx .045$  and  $\text{argsup}_{r \in [1, \infty]} \text{HLAE}_J^S(r, 2\sqrt{3}/7) \approx 1.3288$  with the supremum  $\approx .9844$ . Furthermore, we observe that  $\text{HLAE}_J^S(r, 2\sqrt{3}/7) > \text{HLAE}_J^S(r, \sqrt{3}/4) > \text{HLAE}_J^S(r, \sqrt{3}/8)$  at each  $r$ . Based on the HLAE analysis for the given  $\mathcal{Y}$  we suggest moderate  $r$  values for moderate segregation, and small  $r$  values for severe segregation.

The explicit form of  $\text{HLAE}_J^A(r, \epsilon)$  is similar which implies  $\text{HLAE}_J^A(r, \epsilon = 0) = 0$  and  $\lim_{r \rightarrow \infty} \text{HLAE}_J^A(r, \epsilon) = 0$ .

We calculate HLAE of  $\rho_n(r, J)$  under  $H_\epsilon^A$  for  $\epsilon = \sqrt{3}/21$ ,  $\epsilon = \sqrt{3}/12$ , and  $\epsilon = 5\sqrt{3}/24$ . In Figure 35 we present  $\text{HLAE}_J^S(r, \epsilon)$  for these  $\epsilon$  values conditional on the realization of  $\mathcal{Y}$  in Figure 30

Note that with  $\epsilon = \sqrt{3}/21$ ,  $\text{HLAE}_J^A(r = 1, \sqrt{3}/21) \approx .0009$  and  $\text{argsup}_{r \in [1, \infty]} \text{HLAE}_J^A(r, \sqrt{3}/21) \approx 1.5734$  with the supremum  $\approx .0157$ . With  $\epsilon = \sqrt{3}/12$ ,  $\text{HLAE}_J^A(r = 1, \sqrt{3}/12) \approx .0168$  and  $\text{argsup}_{r \in [1, \infty]} \text{HLAE}_J^A(r, \sqrt{3}/12) \approx 1.6732$  with the supremum  $\approx .1818$ . With  $\epsilon = 5\sqrt{3}/24$ ,  $\text{HLAE}_J^A(r = 1, 5\sqrt{3}/24) \approx .0017$  and  $\text{argsup}_{r \in [1, \infty]} \text{HLAE}_J^A(r, 5\sqrt{3}/24) \approx 3.2396$  with the supremum  $\approx 5.7616$ . Furthermore, we observe that  $\text{HLAE}_J^A(r, 5\sqrt{3}/24) > \text{HLAE}_J^A(r, \sqrt{3}/12) > \text{HLAE}_J^A(r, \sqrt{3}/21)$  at each  $r$ . Based on the HLAE analysis for

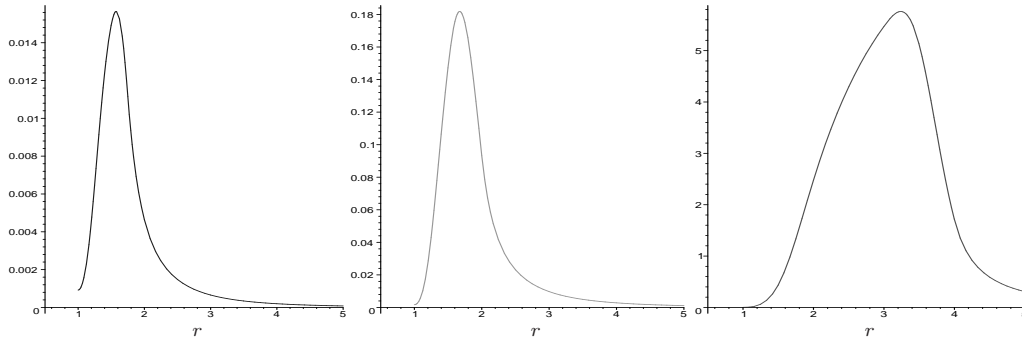


Figure 35: Hodges-Lehmann asymptotic efficacy against association alternative  $H_\epsilon^A$  as a function of  $r$  for  $\epsilon = \sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24$  (left to right) conditional on the realization of  $\mathcal{Y}$  in Figure 29.

the given  $\mathcal{Y}$  we suggest moderate  $r$  values for moderate association and larger  $r$  values for severe association.

## 4 Discussion

The extension to  $\mathbb{R}^d$  for  $d > 2$  is straightforward. See Ceyhan and Priebe (2003b) for more detail. Moreover, the geometry invariance, asymptotic normality of the  $U$ -statistic and consistency of the tests hold for  $d > 2$ .

The first proximity map similar to the  $r$ -factor proximity map  $N_{\mathcal{Y}}^r$  in literature is the spherical proximity map  $N_S(x) := B(x, r(x))$ , (see the references for CCCD in the Introduction). A slight variation of  $N_S$  is the arc-slice proximity map  $N_{AS}(x) := B(x, r(x)) \cap T(x)$  where  $T(x)$  is the Delaunay cell that contains  $x$  (see Ceyhan and Priebe (2003a)). Furthermore, Ceyhan and Priebe introduced the central similarity proximity map  $N_{CS}$  in Ceyhan and Priebe (2003a) and  $N_{\mathcal{Y}}^r$  in Ceyhan and Priebe (2003b). The  $r$ -factor proximity map, when compared to the others, has the advantages that the asymptotic distribution of the domination number  $\gamma_n(N_{\mathcal{Y}}^r)$  is tractable (see Ceyhan and Priebe (2003b)), the exact minimum dominating sets can be found in polynomial time. Moreover  $N_{\mathcal{Y}}^r$  and  $N_{CS}$  are geometry invariant for uniform data over triangles. Additionally, the mean and variance of  $\rho_n$  is not analytically tractable for  $N_S$  and  $N_{AS}$ . While  $N_{\mathcal{Y}}^r(x)$ ,  $N_{CS}(x)$ , and  $N_{AS}(x)$  are well defined only for  $x \in C_H(\mathcal{Y})$ , the convex hull of  $\mathcal{Y}$ ,  $N_S(x)$  is well defined for all  $x \in \mathbb{R}^d$ .  $N_S$  and  $N_{AS}$  require no effort to extend to higher dimensions.

There are many tests available for segregation and association in literature. See Dixon (1994) for a survey on these tests and relevant references. The most prevalent of these tests are Pielou's  $\chi^2$  test of independence and Ripley's test based on  $K(t)$  and  $L(t)$  functions. However, the test we introduce here is not comparable to either of them, since it is a conditional test — conditional on a realization of  $J = |\mathcal{Y}|$  and  $\mathcal{W}$  and we require the number of triangles  $J$  is fixed and relatively small compared to  $n = |\mathcal{X}_n|$ . The null hypothesis for testing spatial patterns has two major forms:

- (i) assuming random labeling of locations, i.e. spatial randomness does not necessarily hold, as in Pielou's test which only tests for the association between classes,
- (ii) assuming not only random labeling but also complete spatial randomness, that is, each class is distributed randomly throughout the area of interest, as in Ripley's test.

Our conditional test is closer to the latter in this regard.

The test based on the mean domination number in Ceyhan and Priebe (2003b) is not a conditional test, but requires both  $n$  and number of Delaunay triangles  $J$  to be large. The comparison for a large but fixed  $J$  is possible. Furthermore, under segregation alternatives, the Pitman asymptotic efficacy is not applicable to the mean domination number case, however, for large  $n$  and  $J$  we suggest the use of it over arc density since for each  $\epsilon > 0$ , Hodges-Lehmann asymptotic efficacy is unbounded for the mean domination number case, while it is bounded for arc density case with  $J > 1$ . As for the association alternative, HLAE suggests moderate  $r$  values which has finite Hodges-Lehmann asymptotic efficacy. So again, for large  $J$  and  $n$  mean domination

number is preferable. The basic advantage of  $\rho_n(r)$  is that, it does not require  $J$  to be large, so for small  $J$  it is preferable.

## Acknowledgments

This work was partially sponsored by the Defense Advanced Research Projects Agency as administered by the Air Force Office of Scientific Research under contract DOD F49620-99-1-0213.

## References

- Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for  $U$ -statistics. *Annals of Statistics*, 6:417–421.
- Ceyhan, E. and Priebe, C. (2003a). Central similarity proximity maps in Delaunay tessellations. In *Proceedings of the Joint Statistical Meeting, Statistical Computing Section, American Statistical Association*.
- Ceyhan, E. and Priebe, C. (2003b). The use of domination number of a random proximity catch digraph for testing segregation/association. Technical Report 642, Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD, 21218. submitted for publication.
- Coomes, D. A., Rees, M., and Turnbull, L. (1999). Identifying aggregation and association in fully mapped spatial data. *Ecology*, 80(2):554–565.
- DeVinney, J., Priebe, C. E., Marchette, D. J., and Socolinsky, D. (2002). Random walks and catch digraphs in classification. <http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf>. Proceedings of the 34<sup>th</sup> Symposium on the Interface: Computing Science and Statistics, Vol. 34.
- Dixon, P. M. (1994). Testing spatial segregation using a nearest-neighbor contingency table. *Ecology*, 75(7):1940–1948.
- Eeden, C. V. (1963). The relation between Pitman’s asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics. *The Annals of Mathematical Statistics*, 34(4):1442–1451.
- Hodges, J. L. J. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the  $t$ -test. *The Annals of Mathematical Statistics*, 27(2):324–335.
- Janson, S., Luczak, T., and Ruciniński, A. (2000). *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York.
- Jaromczyk, J. W. and Toussaint, G. T. (1992). Relative neighborhood graphs and their relatives. *Proceedings of IEEE*, 80:1502–1517.
- Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics, Volume 2, 4th edition*. Griffin, London.
- Lehmann, E. L. (1988). *Nonparametrics: Statistical Methods Based on Ranks*. Prentice-Hall, Upper Saddle River, NJ.
- Marchette, D. J. and Priebe, C. E. (2003). Characterizing the scale dimension of a high dimensional classification problem. *Pattern Recognition*, 36(1):45–60.
- Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley.
- Priebe, C. E., DeVinney, J. G., and Marchette, D. J. (2001). On the distribution of the domination number of random class cover catch digraphs. *Statistics & Probability Letters*, 55:239–246.
- Priebe, C. E., Marchette, D. J., DeVinney, J., and Socolinsky, D. (2003a). Classification using class cover catch digraphs. *Journal of Classification*, 20(1):3–23.

Priebe, C. E., Solka, J. L., Marchette, D. J., and Clark, B. T. (2003b). Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays. *Computational Statistics & Data Analysis on Visualization*, 43-4:621–632.

Toussaint, G. T. (1980). The relative neighborhood graph of a finite planar set. *Pattern Recognition*, 12(4):261–268.

## Appendix 1: Derivation of $\mu(r)$ and $\nu(r)$

In the standard equilateral triangle, let  $y_1 = (0, 0)$ ,  $y_2 = (1, 0)$ ,  $y_3 = (1/2, \sqrt{3}/2)$ ,  $M_C$  be the center of mass,  $M_j$  be the midpoints of the edges  $e_j$  for  $j = 1, 2, 3$ . Then  $M_C = (1/2, \sqrt{3}/6)$ ,  $M_1 = (3/4, \sqrt{3}/4)$ ,  $M_2 = (1/4, \sqrt{3}/4)$ ,  $M_3 = (1/2, 0)$ .

Recall that  $\mathbf{E}[\rho_n(r)] = \frac{1}{n(n-1)} \sum \sum_{i < j} \mathbf{E}[h_{ij}] = \frac{1}{2} \mathbf{E}[h_{12}] = \mu(r) = P(X_j \in N_{\mathcal{Y}}^r(X_i))$ .

Let  $\mathcal{X}_n$  be a random sample of size  $n$  from  $\mathcal{U}(T(\mathcal{Y}))$ . For  $x_1 = (u, v)$ ,  $\ell_r(x_1) = r v + r \sqrt{3} u - \sqrt{3} x$ . Next, let  $N_1 := \ell_r(x_1) \cap e_3$  and  $N_2 := \ell_r(x_1) \cap e_2$ . Then for  $z_1 \in T_s := T(y_1, M_3, M_C)$ ,  $N_{\mathcal{Y}}^r(z_1) = T(y_1, N_1, N_2)$  provided that  $\ell_r(x_1)$  is not outside of  $T(\mathcal{Y})$ , where

$$N_1 = (r(y_1 + \sqrt{3}x_1)\sqrt{3}/3, 0) \text{ and } N_2 = (r(y_1 + \sqrt{3}x_1)\sqrt{3}/6, (y_1 + \sqrt{3}x_1)r/2).$$

### Derivation of $\mu(r)$ in Theorem 2

Now we find  $\mu(r)$  for  $r \in [1, \infty)$ . Observe that, by symmetry,

$$\mu(r) = P(X_2 \in N_{\mathcal{Y}}^r(X_1)) = 6 P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_1 \in T_s).$$

Let  $\ell_s(r, x)$  be the line such that  $r d(y_1, \ell_s(r, x)) = d(y_1, e_1)$ , so  $\ell_s(r, x) = \sqrt{3}(1/r - x)$ . Then if  $x_1 \in T_s$  is above  $\ell_s(r, x)$  then  $N_{\mathcal{Y}}^r(x_1) = T(\mathcal{Y})$ , otherwise,  $N_{\mathcal{Y}}^r(x_1) \subsetneq T(\mathcal{Y})$ .

For  $r \in [1, 3/2)$ ,  $\ell_s(r, x) \cap T_s = \emptyset$ , so  $N_{\mathcal{Y}}^r(x) \subsetneq T(\mathcal{Y})$  for all  $x \in T_s$ . Then

$$\mu(r) = 6 P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_1 \in T_s) = 6 \int_0^{1/2} \int_0^{x/\sqrt{3}} \frac{A(N_{\mathcal{Y}}^r(x_1))}{A(T(\mathcal{Y}))^2} dy dx = 6 \left( \frac{37}{1296} r^2 \right) = \frac{37}{216} r^2.$$

where  $A(N_{\mathcal{Y}}^r(x_1)) = \frac{\sqrt{3}}{12} r^2 (y + \sqrt{3}x)^2$  and  $A(T(\mathcal{Y})) = \sqrt{3}/4$ .

For  $r \in [3/2, 2)$ ,  $\ell_s(r, x)$  crosses through  $\overline{M_3 M_C}$ . Let the  $x$  coordinate of  $\ell_s(r, x) \cap \overline{y_1 M_C}$  be  $s_1$ , then  $s_1 = \frac{3}{4r}$ . See Figure 36.

Then

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_1 \in T_s) &= \int_0^{s_1} \int_0^{x/\sqrt{3}} \frac{A(N_{\mathcal{Y}}^r(x_1))}{A(T(\mathcal{Y}))^2} dy dx + \int_{s_1}^{1/2} \int_0^{\ell_s(r, x)} \frac{A(N_{\mathcal{Y}}^r(x_1))}{A(T(\mathcal{Y}))^2} dy dx + \\ &\int_{s_1}^{1/2} \int_{\ell_s(r, x)}^{x/\sqrt{3}} \frac{1}{A(T(\mathcal{Y}))} dy dx = -\frac{-36 + r^4 + 64r - 32r^2}{48r^2}. \end{aligned}$$

Hence for  $r \in [3/2, 2)$ ,  $\mu(r) = -\frac{1}{8} r^2 - 8r^{-1} + \frac{9}{2} r^{-2} + 4$ .

For  $r \in [2, \infty)$ ,  $\ell_s(r, x)$  crosses through  $\overline{y_1 M_3}$ . Let the  $x$  coordinate of  $\ell_s(r, x) \cap \overline{y_1 M_3}$  be  $s_2$ , then  $s_2 = 1/r$ . See Figure 36.

Then

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_1 \in T_s) &= \int_0^{s_1} \int_0^{x/\sqrt{3}} \frac{A(N_{\mathcal{Y}}^r(x_1))}{A(T(\mathcal{Y}))^2} dy dx + \int_{s_1}^{s_2} \int_0^{\ell_s(r, x)} \frac{A(N_{\mathcal{Y}}^r(x_1))}{A(T(\mathcal{Y}))^2} dy dx \\ &+ \int_{s_1}^{s_2} \int_{\ell_s(r, x)}^{x/\sqrt{3}} \frac{1}{A(T(\mathcal{Y}))} dy dx + \int_{s_2}^{1/2} \int_0^{x/\sqrt{3}} \frac{1}{A(T(\mathcal{Y}))} dy dx = \frac{3 + 2r^2}{12r^2}. \end{aligned}$$

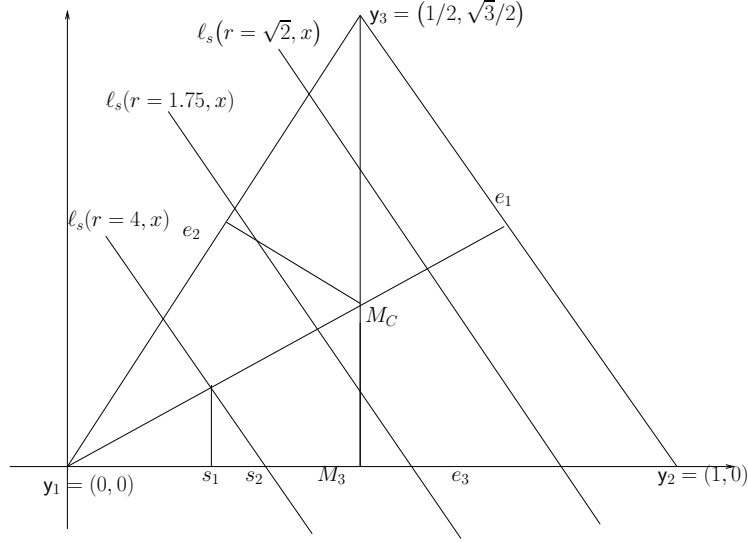


Figure 36: The cases for relative position of  $\ell_s(r, x)$  with various  $r$  values.

Hence for  $r \in [2, \infty)$ ,  $\mu(r) = 1 - \frac{3}{2}r^{-2}$ .

For  $r = \infty$ ,  $\mu(r) = 1$  follows trivially.

## Derivation of $\nu(r)$ in Theorem 2

To find  $\mathbf{Cov}[h_{12}, h_{13}]$ , we introduce a related concept.

**Definition:** Let  $(\Omega, \mathcal{M})$  be a measurable space and consider the proximity map  $N : \Omega \times \wp(\Omega) \rightarrow \wp(\Omega)$ , where  $\wp(\cdot)$  represents the power set functional. For  $B \subset \Omega$ , the  $\Gamma_1$ -region,  $\Gamma_1(\cdot) = \Gamma_1(\cdot, N) : \Omega \rightarrow \wp(\Omega)$  associates the region  $\Gamma_1(B) := \{z \in \Omega : B \subseteq N(z)\}$  with each set  $B \subset \Omega$ . For  $x \in \Omega$ , we denote  $\Gamma_1(\{x\})$  as  $\Gamma_1(x)$ . Note that  $\Gamma_1$ -region depends on proximity region  $N(\cdot)$ .

Furthermore, let  $\Gamma_1(\cdot, N_Y^r)$  be the  $\Gamma_1$ -region associated with  $N_Y^r(\cdot)$ , let  $A_{ij}$  be the event that  $\{X_i X_j \in \mathcal{A}\} = \{X_i \in N_Y^r(X_j)\}$ , then  $h_{ij} = I(A_{ij}) + I(A_{ji})$ . Let

$$P_{2N}^r := P(\{X_2, X_3\} \subset N_Y^r(X_1)), \quad P_M^r := P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r)), \quad P_{2G}^r := P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r)).$$

Then  $\mathbf{Cov}[h_{12}, h_{13}] = \mathbf{E}[h_{12} h_{13}] - \mathbf{E}[h_{12}]\mathbf{E}[h_{13}]$  where

$$\begin{aligned} \mathbf{E}[h_{12} h_{13}] &= \mathbf{E}[(I(A_{12}) + I(A_{21}))(I(A_{13}) + I(A_{31}))] \\ &= P(A_{12} \cap A_{13}) + P(A_{12} \cap A_{31}) + P(A_{21} \cap A_{13}) + P(A_{21} \cap A_{31}). \\ &= P(\{X_2, X_3\} \subset N_Y^r(X_1)) + 2P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r)) + P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r)) \\ &= P_{2N}^r + 2P_M^r + P_{2G}^r. \end{aligned}$$

$$\text{So } \nu(r) = \mathbf{Cov}[h_{12}, h_{13}] = (P_{2N}^r + 2P_M^r + P_{2G}^r) - [2\mu(r)]^2.$$

Furthermore, for any  $x_1 = (u, v) \in T(\mathcal{Y})$ ,  $\Gamma_1(x_1, N_Y^r)$  is a convex or nonconvex polygon. Let  $\xi_j(r, x)$  be the line between  $x_1$  and the vertex  $y_j$  parallel to the edge  $e_j$  such that  $r d(y_j, \xi_j(r, x)) = d(y_j, \ell_r(x_1))$  for  $j = 1, 2, 3$ . Then  $\Gamma_1(x_1, N_Y^r) \cap R(y_j)$  is bounded by  $\xi_j(r, x)$  and the median lines.

For  $x_1 = (u, v)$ ,  $\xi_1(r, x) = -\sqrt{3}x + (v + \sqrt{3}u)/r$ ,  $\xi_2(r, x) = (v + \sqrt{3}r(x-1) + \sqrt{3}(1-u))/r$  and  $\xi_3(r, x) = (\sqrt{3}(r-1) + 2v)/(2r)$ . To find the covariance, we need to find the possible types of  $\Gamma_1(x_1, N_Y^r)$  and  $N_Y^r(x_1)$  for  $r \in [1, \infty)$ .

We partition  $[1, \infty)$  with respect to the types of  $N_Y^r(x_1)$  and  $\Gamma_1(x_1, N_Y^r)$  and obtain  $[1, 4/3)$ ,  $[4/3, 3/2)$ ,  $[3/2, 2)$ ,  $[2, \infty)$ .

For  $r \in [1, 4/3)$ , there are six cases regarding  $\Gamma_1(x_1, N_Y^r)$  and one case for  $N_Y^r(x_1)$ . See Figure 37 for the prototypes of these six cases of  $\Gamma_1(x_1, N_Y^r)$ . Each case  $j$ , corresponds to the region  $R_j$  in Figure 38, where

$\ell_{am}(x) = x/\sqrt{3}$ ,  $q_1(x) = (2r + 3x - 3)/\sqrt{3}$ ,  $q_2(x) = \sqrt{3}(1/2 - r/3)$ ,  $q_3(x) = \sqrt{3}(x - 1 + r/2)$ ,  $q_4(x) = \sqrt{3}(1/2 - r/4)$ ,  $q_{12}(x) = \sqrt{3}(r/2 - x)$  and  $s_1 = 1 - 2r/3$ ,  $s_2 = 3/2 - r$ ,  $s_3 = 1 - r/2$ ,  $s_4 = 3/2 - 5r/6$ ,  $s_5 = 3r/8$ . The explicit forms of  $R_j$ ,  $j = 1, \dots, 6$  are as follows:

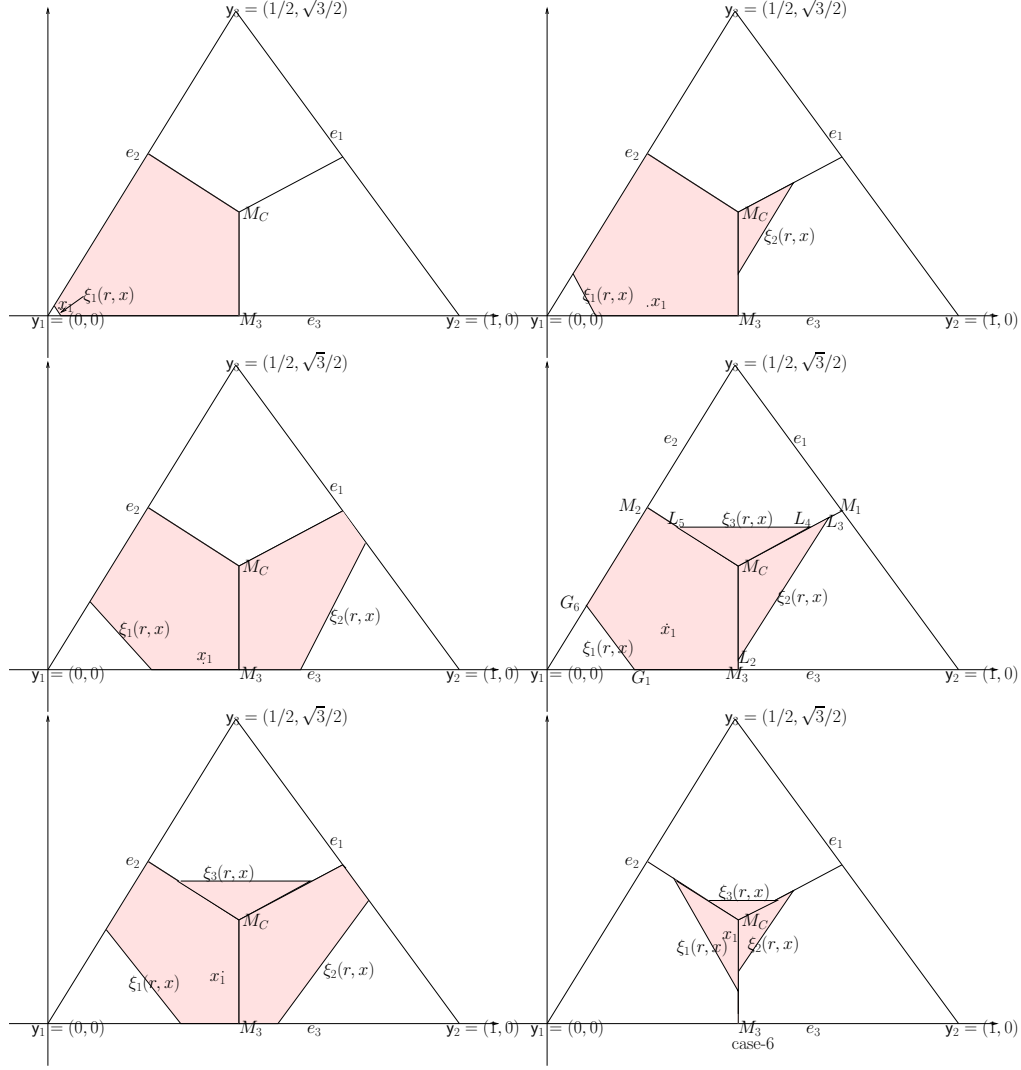


Figure 37: The prototypes of the six cases of  $\Gamma_1(x_1, N_Y^r)$  for  $x_1 \in T_s$  for  $r \in [1, 4/3]$ .

$$\begin{aligned}
R_1 &= \{(x, y) \in [0, s_1] \times [0, \ell_{am}(x)] \cup [s_1, s_2] \times [q_1(x), \ell_{am}(x)]\}, \\
R_2 &= \{(x, y) \in [s_1, s_2] \times [0, q_1(x)] \cup [s_2, s_3] \times [0, q_2(x)] \cup [s_3, s_4] \times [q_3(x), q_2(x)]\}, \\
R_3 &= \{(x, y) \in [s_3, s_4] \times [0, q_3(x)] \cup [s_4, 1/2] \times [0, q_2(x)]\}, \\
R_4 &= \{(x, y) \in [s_1, s_2] \times [0, q_1(x)] \cup [s_4, s_5] \times [q_3(x), \ell_{am}(x)] \cup [s_5, 1/2] \times [q_3(x), q_{12}(x)]\}, \\
R_5 &= \{(x, y) \in [s_4, 1/2] \times [q_2(x), q_3(x)]\}, \quad R_6 = \{(x, y) \in [s_5, 1/2] \times [q_{12}(x), \ell_{am}(x)]\}.
\end{aligned}$$

By symmetry,  $P_{2N}^r = 6 P(\{X_2, X_3\} \subset N_Y^r(X_1), X_1 \in T_s)$ .

For  $r \in [1, 4/3]$ ,

$$P(\{X_2, X_3\} \subset N_Y^r(X_1), X_1 \in T_s) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_Y^r(x_1))^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{781}{116640} r^4,$$

where  $A(N_Y^r(x_1)) = \frac{\sqrt{3}}{12} r^2 (y + \sqrt{3}x)^2$ . Hence for  $r \in [1, 4/3]$ ,  $P_{2N}^r = \frac{781}{19440} r^4$ . Note that the same results also hold for  $r \in [4/3, 3/2]$ .



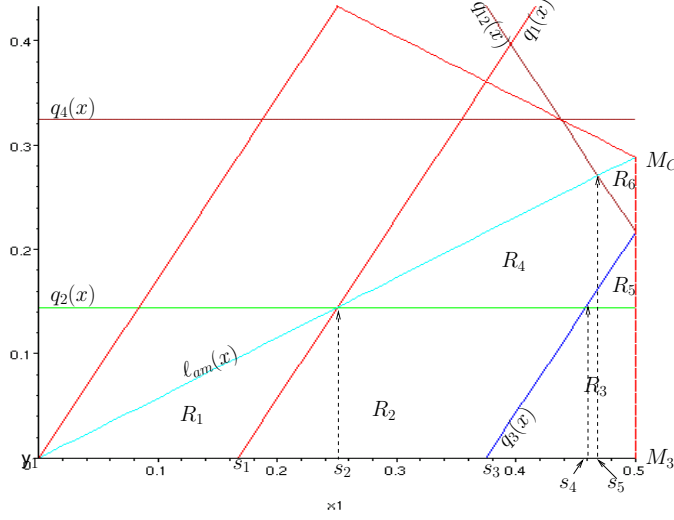


Figure 38: The regions corresponding to the prototypes of the six cases for  $r \in [1, 4/3)$  with  $r = 1.25$ .

Next, by symmetry,  $P_{2G}^r = 6 P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in T_s)$ , and

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in T_s) = \sum_{j=1}^6 P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_j).$$

For  $x_1 \in R_1$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_1) &= \int_0^{s_1} \int_0^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_1}^{s_2} \int_{q_1(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{(211r^4 - 1716r^3 + 5751r^2 - 6696r + 2511)(2r - 3)^2}{10935r^4}, \end{aligned}$$

where  $A(\Gamma_1(x_1, N_{\mathcal{Y}}^r)) = -\frac{\sqrt{3}(r^2 - (\sqrt{3}x + y)^2)}{12r^2}$ .

For  $x_1 \in R_2$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_2) &= \int_{s_1}^{s_2} \int_0^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_2}^{s_3} \int_0^{q_2(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_3}^{s_4} \int_{q_3(x)}^{q_2(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &= -\frac{(2r - 3)(440r^4 - 4091r^3 + 13476r^2 - 16506r + 6696)}{9720r^3}. \end{aligned}$$

where  $A(\Gamma_1(x_1, N_{\mathcal{Y}}^r)) = \frac{\sqrt{3}(-4\sqrt{3}ry - 12r + 12rx + 5r^2 + 2y^2 + 6\sqrt{3}y - 8x\sqrt{3}y + 9 - 18x + 6x^2)}{12r^2}$ .

For  $x_1 \in R_3$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_3) &= \int_{s_3}^{s_4} \int_0^{q_3(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_4}^{1/2} \int_0^{q_2(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &= -\frac{(2r - 3)(21056r^5 - 7845r^4 + 231300r^3 - 943650r^2 + 1127520r - 428652)}{262440r^4}. \end{aligned}$$

where  $A(\Gamma_1(x_1, N_{\mathcal{Y}}^r)) = -\frac{\sqrt{3}(2y^2 + 2\sqrt{3}y + 3 - 6x + 6x^2 - 2r^2)}{12r^2}$ .

For  $x_1 \in R_4$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in R_4) &= \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_4}^{s_5} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_5}^{1/2} \int_{q_3(x)}^{q_{12}(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &= -\frac{12873091}{699840} r^2 + \frac{81239}{648} r + \frac{14714}{27} r^{-1} - \frac{4238}{9} r^{-2} + \frac{656}{3} r^{-3} - \frac{128}{3} r^{-4} - \frac{77123}{216}. \end{aligned}$$

$$\text{where } A(\Gamma_1(x_1, N_Y^r)) = \frac{\sqrt{3}(9r^2 + 18 - 24r + 4\sqrt{3}ry - 18x + 6x^2 + 14y^2 + 12rx - 8x\sqrt{3}y - 6\sqrt{3}y)}{12r^2}.$$

For  $x_1 \in R_5$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in R_5) &= \int_{s_4}^{1/2} \int_{q_2(x)}^{q_3(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &= \frac{(89305r^4 - 364080r^3 + 598320r^2 - 468288r + 145152)(-6 + 5r)^2}{262440r^4}. \end{aligned}$$

$$\text{where } A(\Gamma_1(x_1, N_Y^r)) = \frac{\sqrt{3}(9r^2 + 18 - 24r + 4\sqrt{3}ry - 18x + 6x^2 + 14y^2 + 12rx - 8x\sqrt{3}y - 6\sqrt{3}y)}{12r^2}.$$

For  $x_1 \in R_6$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in R_6) &= \int_{s_5}^{1/2} \int_{q_{12}(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &= \frac{(1081r^4 - 4672r^3 + 7624r^2 - 5568r + 1536)(-4 + 3r)^2}{960r^4}. \end{aligned}$$

$$\text{where } A(\Gamma_1(x_1, N_Y^r)) = -\frac{\sqrt{3}(\sqrt{3}y - 2r^2 - 3 + 3x + 4r - 3x^2 - 3y^2)}{2r^2}.$$

So

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r)) &= 6 \left( \frac{25687}{349920} r^2 - \frac{133}{972} r + \frac{14}{81} r^{-1} - \frac{1}{9} r^{-2} + \frac{1}{90} r^{-4} - \frac{1}{324} \right) \\ &= \frac{25687r^6 - 47880r^5 - 1080r^4 + 60480r^3 - 38880r^2 + 3888}{58320r^4}. \end{aligned}$$

Furthermore, by symmetry,  $P_M^r = 6P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in T_s)$ , and

$$P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in T_s) = \sum_{j=1}^6 P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in R_j).$$

For  $x_1 \in R_1$ ,

$$\begin{aligned} P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in R_1) &= \int_0^{s_1} \int_0^{\ell_{am}(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_1}^{s_2} \int_{q_1(x)}^{\ell_{am}(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx = -\frac{1}{21870} (143r^2 - 744r + 558)(2r - 3)^4. \end{aligned}$$

For  $x_1 \in R_2$ ,

$$\begin{aligned} P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in R_2) &= \int_{s_1}^{s_2} \int_0^{\ell_{am}(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_2}^{s_3} \int_0^{q_2(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_3}^{s_4} \int_{q_3(x)}^{q_2(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx \\ &= \frac{1}{349920} r(2r - 3)(23014r^4 - 187311r^3 + 517896r^2 - 594216r + 241056). \end{aligned}$$

For  $x_1 \in R_3$ ,

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_3) \\ = \int_{s_3}^{s_4} \int_0^{q_3(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_4}^{1/2} \int_0^{q_2(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ = \frac{1}{1049760} (2r - 3)(874r^5 - 297327r^4 + 1858392r^3 - 4298832r^2 + 4280202r - 1546209). \end{aligned}$$

For  $x_1 \in R_4$ ,

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_4) = \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ + \int_{s_4}^{s_5} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_5}^{1/2} \int_{q_3(x)}^{q_{12}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ = -\frac{1}{466560} r (1762560r - 497664 - 2661120r^2 + 201395r^5 - 1017720r^4 + 2212560r^3). \end{aligned}$$

For  $x_1 \in R_5$ ,

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_5) = \int_{s_4}^{1/2} \int_{q_2(x)}^{q_3(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ = \frac{1}{262440} (1570r^4 - 1380r^3 - 11205r^2 + 29700r - 19116)(-6 + 5r)^2. \end{aligned}$$

For  $x_1 \in R_6$ ,

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_6) = \int_{s_5}^{1/2} \int_{q_{12}(x)}^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ = \frac{1}{51840} (1485r^4 - 2064r^3 + 16r^2 - 128r + 768)(-4 + 3r)^2. \end{aligned}$$

Thus

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r)) &= 6 \left( \frac{3007}{699840} r^6 - \frac{8}{405} r^5 + \frac{5}{648} r^4 + \frac{1}{9} r^3 - \frac{133}{648} r^2 + \frac{56}{405} r - \frac{143}{4320} \right) \\ &= \frac{3007}{116640} r^6 - \frac{16}{135} r^5 + \frac{5}{108} r^4 + \frac{2}{3} r^3 - \frac{133}{108} r^2 + \frac{112}{135} r - \frac{143}{720}. \end{aligned}$$

Hence

$$\begin{aligned} \mathbf{E}[h_{12} h_{13}] &= \left[ 3007r^{10} - 13824r^9 + 7743r^8 + 77760r^7 - 117953r^6 + 48888r^5 - 24246r^4 + 60480r^3 \right. \\ &\quad \left. - 38880r^2 + 3888 \right] / \left[ 58320r^4 \right]. \end{aligned}$$

Thus

$$\begin{aligned} \nu(r) &= \left[ 3007r^{10} - 13824r^9 + 898r^8 + 77760r^7 - 117953r^6 + 48888r^5 - 24246r^4 + 60480r^3 - 38880r^2 \right. \\ &\quad \left. + 3888 \right] / \left[ 58320r^4 \right]. \end{aligned}$$

For  $r \in [4/3, 3/2]$ , there are six cases regarding  $\Gamma_1(x_1, N_{\mathcal{Y}}^r)$  and one case for  $N_{\mathcal{Y}}^r(x_1)$ . Prototypes of the five of the cases for  $\Gamma_1(x_1, N_{\mathcal{Y}}^r)$  are as in case- $j$  for  $j = 1, \dots, 5$  in Figure 37 and the new case, case-7, is depicted in Figure 39. Each case  $j$  corresponds to the region  $R_j$  in Figure 40 where  $s_1 = 1 - 2r/3$ ,  $s_2 = 3/2 - r$ ,  $s_3 = 1 - r/2$ ,  $s_4 = 3/2 - 5r/6$ ,  $s_5 = 3/2 - 3r/4$ . The explicit forms of  $R_j$ ,  $j = 1, 2, 3$  are same as before, for  $j = 4, 5, 7$  are given below:

$$\begin{aligned} R_4 &= \{(x, y) \in [s_2, s_4] \times [q_2(x), \ell_{am}(x)] \cup [s_4, s_6] \times [q_3(x), \ell_{am}(x)]\} \\ R_5 &= \{(x, y) \in [s_4, s_6] \times [q_2(x), q_3(x)] \cup [s_6, 1/2] \times [q_2(x), q_4(x)]\} \\ R_7 &= \{(x, y) \in [s_6, 1/2] \times [q_4(x), \ell_{am}(x)]\} \end{aligned}$$

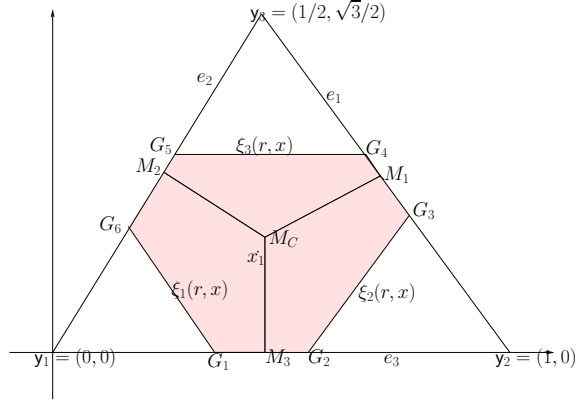


Figure 39: The prototype of the new case for  $\Gamma_1(x_1, N_Y^r)$  for  $x_1 \in T_s$  for  $r \in [4/3, 3/2]$ .

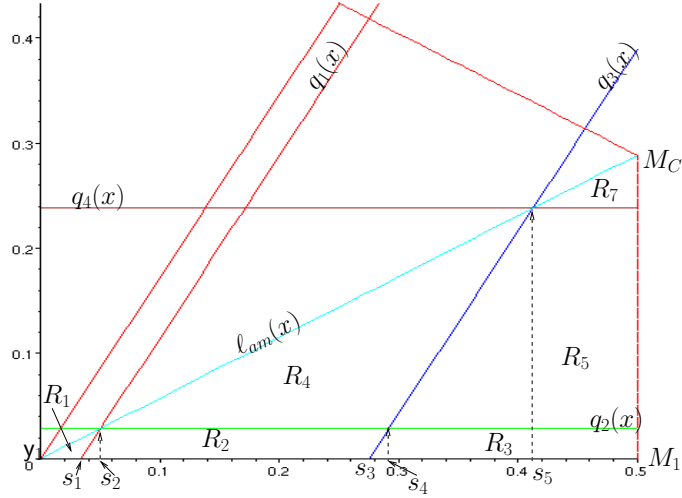


Figure 40: The regions corresponding to the six cases for  $r \in [4/3, 3/2]$

where  $\ell_{am}(x) = x/\sqrt{3}$ ,  $q_1(x) = (2r-3)/\sqrt{3} + \sqrt{3}x$ ,  $q_2(x) = \sqrt{3}(1/2 - r/3)$ ,  $q_3(x) = \sqrt{3}(x - 1 + r/2)$ , and  $q_4(x) = \sqrt{3}(1/2 - r/4)$ .

Then  $P_{2N}^r = \frac{781r^4}{19440}$ . We use the same limits of integration in  $\mu(r)$  calculations with the integrand  $A(N_Y^r(x_1))^2/A(T(\mathcal{Y}))^3$ .

Next, by symmetry,  $P_{2G}^r = 6P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in T_s)$ , and, let  $S_I := \{1, 2, 3, 4, 5, 7\}$ , then

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in T_s) = \sum_{j=S_I} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in R_j).$$

For  $x_1 \in R_j$ ,  $j = 1, 2, 3$  we get the same result as before.

For  $x_1 \in R_4$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in R_4) &= \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_4}^{s_6} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{9637r^4 - 89640r^3 + 288360r^2 - 362880r + 155520}{349920r^2}. \end{aligned}$$

where  $A(\Gamma_1(x_1, N_Y^r)) = \frac{\sqrt{3}(9r^2 + 18 - 24r + 4\sqrt{3}ry - 18x + 6x^2 + 14y^2 + 12rx - 8x\sqrt{3}y - 6\sqrt{3}y)}{12r^2}$ .

For  $x_1 \in R_5$ ,

$$\begin{aligned}
& P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in R_5) \\
&= \int_{s_4}^{s_6} \int_{q_2(x)}^{q_3(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_6}^{1/2} \int_{q_2(x)}^{q_4(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\
&= \frac{87251 r^5 + 13219200 r - 11214720 r^2 - 5225472 + 3377160 r^3 - 261288 r^4}{2099520 r^3}.
\end{aligned}$$

where  $A(\Gamma_1(x_1, N_Y^r))$  is same as before.

For  $x_1 \in R_7$ ,

$$\begin{aligned}
P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_Y^r), X_1 \in R_7) &= \int_{s_6}^{1/2} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_Y^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\
&= \frac{(57 r^4 + 96 r^3 - 72 r^2 - 576 r + 512)(-4 + 3 r)^2}{2880 r^4}.
\end{aligned}$$

where  $A(\Gamma_1(x_1, N_Y^r)) = -\frac{\sqrt{3}(6 y^2 - 2\sqrt{3} y + 6 - 6 x + 6 x^2 - 3 r^2)}{12 r^2}$ .

So,

$$\begin{aligned}
P_{2G}^r &= 6 \left( \frac{-47880 r^5 - 38880 r^2 + 25687 r^6 - 1080 r^4 + 60480 r^3 + 3888}{349920 r^4} \right) \\
&= \frac{-47880 r^5 - 38880 r^2 + 25687 r^6 - 1080 r^4 + 60480 r^3 + 3888}{58320 r^4}.
\end{aligned}$$

Furthermore,

$$P_M^r = \sum_{j \in S_I} P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in R_j).$$

For  $x_1 \in R_j$ ,  $j = 1, 2, 3$  we get the same result as before.

For  $x_1 \in R_4$ ,

$$\begin{aligned}
& P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in R_4) \\
&= \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_4}^{s_6} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx \\
&= -\frac{1}{466560} r^2 (207360 + 404640 r^2 - 483840 r - 142920 r^3 + 17687 r^4).
\end{aligned}$$

For  $x_1 \in R_5$ ,

$$\begin{aligned}
& P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in R_5) \\
&= \int_{s_4}^{s_6} \int_{q_2(x)}^{q_3(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_6}^{1/2} \int_{q_2(x)}^{q_4(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx \\
&= -\frac{r(399064320 r - 150792192 + 171990000 r^3 - 391461120 r^2 - 31140648 r^4 + 1230359 r^5)}{67184640}.
\end{aligned}$$

For  $x_1 \in R_7$ ,

$$\begin{aligned}
P(X_2 \in N_Y^r(X_1), X_3 \in \Gamma_1(X_1, N_Y^r), X_1 \in R_7) &= \int_{s_6}^{1/2} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A(N_Y^r(x_1)) A(\Gamma_1(x_1, N_Y^r))}{A(T(\mathcal{Y}))^3} dy dx \\
&= \frac{1}{829440} (2727 r^4 - 3648 r^3 - 52736 r^2 + 166656 r - 121600)(-4 + 3 r)^2.
\end{aligned}$$

Then,

$$P_M^r = 6 \left( \frac{5467}{2799360} r^6 - \frac{35}{2592} r^5 + \frac{37}{1296} r^4 - \frac{13}{648} r^2 + \frac{83}{12960} \right) = \frac{5467}{466560} r^6 - \frac{35}{432} r^5 + \frac{37}{216} r^4 - \frac{13}{108} r^2 + \frac{83}{216}.$$

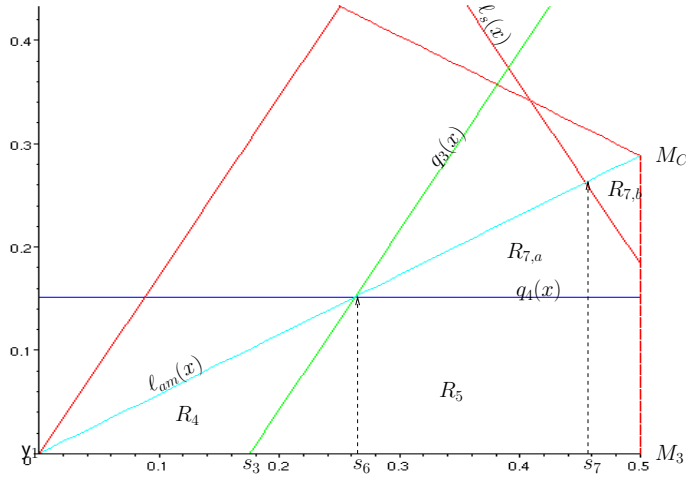


Figure 41: The regions corresponding to the three cases for  $r \in [3/2, 2)$  with  $r = 1.65$

So,

$$\mathbf{E}[h_{12} h_{13}] = \left[ 5467 r^{10} - 37800 r^9 + 89292 r^8 + 46588 r^6 - 191520 r^5 + 13608 r^4 + 241920 r^3 - 155520 r^2 + 15552 \right] / \left[ 233280 r^4 \right].$$

Thus, for  $r \in [4/3, 3/2)$

$$\nu(r) = \left[ 5467 r^{10} - 37800 r^9 + 61912 r^8 + 46588 r^6 - 191520 r^5 + 13608 r^4 + 241920 r^3 - 155520 r^2 + 15552 \right] / \left[ 233280 r^4 \right].$$

For  $r \in [3/2, 2)$ , there are three cases regarding  $\Gamma_1(x_1, N_{\mathcal{Y}}^r)$  and two cases for  $N_{\mathcal{Y}}^r(x_1)$ . The prototypes of these three cases as in cases 4, 5, and 7 of Figures 37 and 39. Each case  $j$ , corresponds to the region  $R_j$  in Figure 41 where  $q_j(x)$  are same as before for  $j = 3, 4$ , and  $s_j$ ,  $j = 3, 4, 6$  are same as before and  $s_7 = 3/(4r)$ . Observe that for  $x_1 \in R_4 \cup R_5 \cup R_{7a}$ ,  $N_{\mathcal{Y}}^r(x_1) = T_r(x_1) \subsetneq T(\mathcal{Y})$ , and for  $x_1 \in R_{7b}$ ,  $N_{\mathcal{Y}}^r(x_1) = T(\mathcal{Y})$ . So there are four regions to consider to calculate the covariance.

Then, for  $x_1 = (x, y) \in R_j$ ,  $\Gamma_1(x_1, N_{\mathcal{Y}}^r)$  are same as before for  $j = 4, 5, 7$ . The explicit forms of  $R_j$ ,  $j = 4, 5, 7a, 7b$  are given below (the explicit form of  $R_7$  is same as before):

$$\begin{aligned} R_4 &= \{(x, y) \in [0, s_3] \times [0, \ell_{am}(x)] \cup [s_3, s_6] \times [q_3(x), \ell_{am}(x)]\} \\ R_5 &= \{(x, y) \in [s_3, s_6] \times [0, q_3(x)] \cup [s_6, 1/2] \times [0, q_4(x)]\} \\ R_{7,a} &= \{(x, y) \in [s_6, s_7] \times [q_4(x), \ell_{am}(x)] \cup [s_7, 1/2] \times [q_4(x), \ell_s(r, x)]\} \\ R_{7,b} &= \{(x, y) \in [s_7, 1/2] \times [\ell_s(r, x), \ell_{am}(x)]\} \end{aligned}$$

Now,

$$\begin{aligned} P(\{X_2, X_3\} \subset N_{\mathcal{Y}}^r(X_1), X_1 \in T_s) &= \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dy dx = \int_0^{s_7} \int_0^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_7}^{1/2} \int_0^{\ell_s(x)} \frac{A(N_{\mathcal{Y}}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_7}^{1/2} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T(\mathcal{Y}))} dy dx = -\frac{480 + r^6 + 768r - 320r^2}{480r^2}. \end{aligned}$$

$$\text{Hence } P_{2N}^r = \frac{-480 + r^6 + 768r - 320r^2}{80r^2}.$$

Next, by symmetry,  $P_{2G}^r = 6 P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in T_s)$ , and, let  $S_I := \{4, 5, 7\}$ , then

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in T_s) = \sum_{j=S_I} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_j).$$

For  $x_1 \in R_4$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_4) &= \int_0^{s_3} \int_0^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_3}^{s_6} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{(237r^4 - 956r^3 + 1728r^2 - 1584r + 592)(-2+r)^2}{480r^4}. \end{aligned}$$

For  $x_1 \in R_5$ ,

$$\begin{aligned} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_5) &= \int_{s_3}^{s_6} \int_0^{q_3(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_6}^{1/2} \int_0^{q_4(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{(r-2)(1909r^5 - 6142r^4 + 10036r^3 - 14808r^2 + 15024r - 6048)}{2880r^4}. \end{aligned}$$

For  $x_1 \in R_7$  the result is same as before. So

$$\begin{aligned} P_{2G}^r &= 6 \left( \frac{7320r^4 - 984r^5 + 13r^6 - 20480r^3 + 27840r^2 - 18816r + 5152}{1440r^4} \right) \\ &= \frac{7320r^4 - 984r^5 + 13r^6 - 20480r^3 + 27840r^2 - 18816r + 5152}{240r^4}. \end{aligned}$$

Furthermore,

$$P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in T_s) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx.$$

For  $x_1 \in R_4$ ,

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_4) &= \int_0^{s_3} \int_0^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_3}^{s_6} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx = \frac{1}{1920} (99r^2 - 16r - 84)(-2+r)^4. \end{aligned}$$

For  $x_1 \in R_5$ ,

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_5) &= \int_{s_3}^{s_6} \int_0^{q_3(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_6}^{1/2} \int_0^{q_4(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ &= -\frac{1}{92160} (-2+r)(7535r^5 - 35210r^4 + 9500r^3 + 181560r^2 - 308880r + 147168). \end{aligned}$$

For  $x_1 \in R_{7a}$ ,

$$\begin{aligned} P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_{7a}) &= \int_{s_6}^{s_7} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ &+ \int_{s_7}^{1/2} \int_{q_4(x)}^{s(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx = \frac{303}{10240} r^6 - \frac{91}{768} r^5 - \frac{53}{128} r^4 + \frac{235}{72} r^3 - \frac{173}{24} r^2 \\ &+ \frac{101}{15} r + \frac{2}{3} r^{-1} - \frac{3}{4} r^{-2} - \frac{16}{9} r^{-3} + 4r^{-4} - \frac{18}{5} r^{-5} + \frac{3}{2} r^{-6} - \frac{34}{15}. \end{aligned}$$

For  $x_1 \in R_{7b}$ ,

$$P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_{7b}) = \int_{s_7}^{1/2} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx = \frac{(2r^4 - 4r^2 + 4r - 3)(2r - 3)^2}{12r^6}.$$

Hence

$$P_{2G}^r = 6 \left( - \left[ 7r^{12} - 72r^{11} + 240r^{10} - 1440r^8 + 3456r^7 - 10296r^6 + 15360r^5 + 6720r^4 - 40960r^3 + 46080r^2 - 27648r + 8640 \right] / \left[ 11520r^6 \right] \right) = - \left[ 7r^{12} - 72r^{11} + 240r^{10} - 1440r^8 + 3456r^7 - 10296r^6 + 15360r^5 + 6720r^4 - 40960r^3 + 46080r^2 - 27648r + 8640 \right] / \left[ 1920r^6 \right].$$

Thus

$$\mathbf{E}[h_{12} h_{13}] = - \left[ 7r^{12} - 72r^{11} + 252r^{10} - 1492r^8 + 7392r^7 - 43416r^6 + 106496r^5 - 110400r^4 + 34304r^3 + 25472r^2 - 27648r + 8640 \right] / \left[ 960r^6 \right].$$

Therefore, for  $r \in [3/2, 2)$

$$\nu(r) = - \left[ 7r^{12} - 72r^{11} + 312r^{10} - 5332r^8 + 15072r^7 + 13704r^6 - 139264r^5 + 273600r^4 - 242176r^3 + 103232r^2 - 27648r + 8640 \right] / \left[ 960r^6 \right].$$

For  $r \in [2, \infty)$ , there is only one case regarding  $\Gamma_1(x_1, N_{\mathcal{Y}}^r)$ , namely  $R_7$ , and two cases regarding  $N_{\mathcal{Y}}^r(x_1)$ . Furthermore,  $s_7$ , is same as before and  $s_8 = 1/r$ . Observe that for  $x_1 \in R_{7a}$ ,  $N_{\mathcal{Y}}^r(x_1) = T_r(x_1) \subsetneq T(\mathcal{Y})$ , and for  $x_1 \in R_{7b}$ ,  $N_{\mathcal{Y}}^r(x_1) = T(\mathcal{Y})$ . So there are two regions to consider to calculate the covariance.

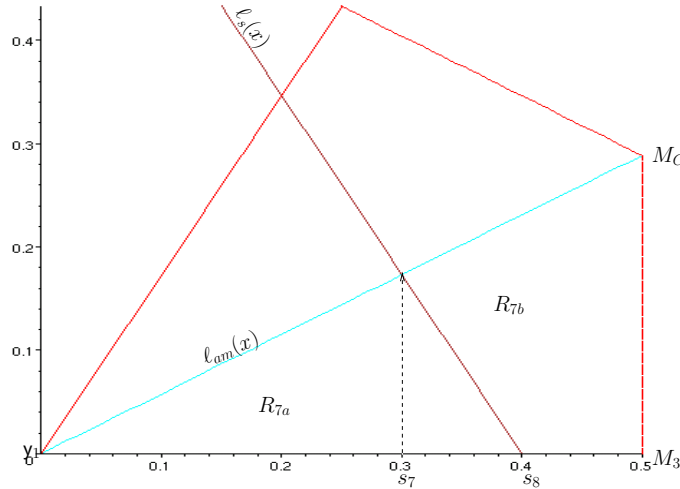


Figure 42: The regions corresponding to the two cases for  $N_{\mathcal{Y}}^r(x_1)$  for  $r \in [2, \infty)$  with  $r = 2.5$

For  $x_1 = (x, y) \in R_7$ ,  $\Gamma_1(x_1, N_{\mathcal{Y}}^r)$  is same as before. The explicit form of  $R_7$ , is same as  $T_s$ . For  $R_{7a}$  and  $R_{7b}$ , see below:

$$R_{7,a} = \{(x, y) \in [0, s_7] \times [0, \ell_{am}(x)] \cup [s_7, s_8] \times [0, \ell_s(r, x)]\}$$

$$R_{7,b} = \{(x, y) \in [s_7, s_8] \times [\ell_s(r, x), \ell_{am}(x)] \cup [s_8, 1/2] \times [0, \ell_{am}(r, x)]\}$$



Now,

$$P(\{X_2, X_3\} \subset N_{\mathcal{Y}}^r(X_1), X_1 \in T_s) = \int_0^{s_7} \int_0^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_7}^{s_8} \int_0^{\ell_s(x)} \frac{A(N_{\mathcal{Y}}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dy dx \\ + \int_{s_7}^{s_8} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T(\mathcal{Y}))} dy dx + \int_{s_8}^{1/2} \int_0^{\ell_{am}(x)} \frac{1}{A(T(\mathcal{Y}))} dy dx = -\frac{1}{3} r^{-2} + \frac{1}{6}.$$

Hence  $P_{2N}^r = 1 - 2r^{-2}$ . Next,

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in T_s) = P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in R_7) \\ = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{34 - 45r^2 + 15r^4}{90r^4}.$$

$$\text{So } P_{2G}^r = \frac{34 - 45r^2 + 15r^4}{15r^4}.$$

Furthermore,

$$P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r), X_1 \in T_s) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ = \int_0^{s_7} \int_0^{\ell_{am}(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx + \int_{s_7}^{s_8} \int_0^{\ell_s(x)} \frac{A(N_{\mathcal{Y}}^r(x_1)) A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^3} dy dx \\ + \int_{s_7}^{s_8} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^2} dy dx + \int_{s_8}^{1/2} \int_0^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r))}{A(T(\mathcal{Y}))^2} dy dx = \frac{25 - 48r + 90r^2 - 90r^4 + 30r^6}{180r^6}.$$

$$\text{So } P_M^r = \frac{25 - 48r + 90r^2 - 90r^4 + 30r^6}{30r^6}.$$

Hence,  $\mathbf{E}[h_{12} h_{13}] = \frac{60r^6 - 165r^4 + 124r^2 - 48r + 25}{15r^6}$ . Thus, for  $r \in [2, \infty)$ ,

$$\nu(r) = \frac{15r^4 - 11r^2 - 48r + 25}{15r^6}.$$

For  $r = \infty$ , it is trivial to see that  $\nu(r) = 0$ .

## Appendix 2: The Mean $\mu(r, \epsilon)$ Under Segregation and Association Alternatives

Derivation of  $\mu(r, \epsilon)$  involves detailed geometric calculations and partitioning of the space of  $(r, \epsilon, x_1)$  for  $r \in [1, \infty)$ ,  $\epsilon \in [0, \sqrt{3}/3)$ , and  $x_1 \in T_s = T_s$ . See Appendix 3 for the derivation of  $\mu(r, \epsilon)$  at a demonstrative interval.

### $\mu_S(r, \epsilon)$ Under Segregation Alternatives

Under segregation, we compute  $\mu_S(r, \epsilon)$  explicitly. For  $\epsilon \in [0, \sqrt{3}/8)$ ,  $\mu_S(r, \epsilon) = \sum_{j=1}^7 \varpi_{1,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$  where

$$\varpi_{1,1}(r, \epsilon) = -\frac{576r^2\epsilon^4 - 1152\epsilon^4 - 37r^2 + 288\epsilon^2}{216(2\epsilon + 1)^2(2\epsilon - 1)^2}, \\ \varpi_{1,2}(r, \epsilon) = -\left[576r^4\epsilon^4 - 1152r^2\epsilon^4 + 91r^4 + 512\sqrt{3}r^3\epsilon + 2592r^2\epsilon^2 + 1536\sqrt{3}r\epsilon^3 + 1152\epsilon^4 \right. \\ \left. - 768r^3 - 2304\sqrt{3}r^2\epsilon - 6912r\epsilon^2 - 2304\sqrt{3}\epsilon^3 + 1728r^2 + 3456\sqrt{3}r\epsilon + 5184\epsilon^2 \right. \\ \left. - 1728r - 1728\sqrt{3}\epsilon + 648\right] / \left[216r^2(2\epsilon + 1)^2(2\epsilon - 1)^2\right],$$

$$\begin{aligned}
\varpi_{1,3}(r, \epsilon) &= - \left[ 192 r^4 \epsilon^4 - 384 r^2 \epsilon^4 + 9 r^4 + 864 r^2 \epsilon^2 + 512 \sqrt{3} r \epsilon^3 + 384 \epsilon^4 - 2304 r \epsilon^2 - 768 \sqrt{3} \epsilon^3 \right. \\
&\quad \left. - 288 r^2 + 1728 \epsilon^2 + 576 r - 324 \right] / \left[ 72 r^2 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2 \right], \\
\varpi_{1,4}(r, \epsilon) &= - \left[ 192 r^4 \epsilon^4 - 384 r^2 \epsilon^4 - 9 r^4 - 96 \sqrt{3} r^3 \epsilon + 288 r^2 \epsilon^2 - 128 \epsilon^4 + 144 r^3 + 576 \sqrt{3} r^2 \epsilon + 256 \right. \\
&\quad \left. \sqrt{3} \epsilon^3 - 720 r^2 - 1152 \sqrt{3} r \epsilon - 576 \epsilon^2 + 1152 r + 768 \sqrt{3} \epsilon - 612 \right] / \left[ 72 r^2 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2 \right], \\
\varpi_{1,5}(r, \epsilon) &= - \frac{48 r^4 \epsilon^4 - 96 r^2 \epsilon^4 + 72 r^2 \epsilon^2 - 32 \epsilon^4 + 64 \sqrt{3} \epsilon^3 - 18 r^2 - 144 \epsilon^2 + 27}{18 r^2 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2}, \\
\varpi_{1,6}(r, \epsilon) &= \frac{48 r^4 \epsilon^4 + 256 r^3 \epsilon^4 - 128 \sqrt{3} r^3 \epsilon^3 + 288 r^2 \epsilon^4 - 192 \sqrt{3} r^2 \epsilon^3 + 72 r^2 \epsilon^2 + 18 r^2 + 48 \sqrt{3} \epsilon - 45}{18 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2 r^2}, \\
\varpi_{1,7}(r, \epsilon) &= 1,
\end{aligned}$$

with the corresponding intervals  $\mathcal{I}_1 = [1, 3/2 - \sqrt{3}\epsilon)$ ,  $\mathcal{I}_2 = [3/2 - \sqrt{3}\epsilon, 3/2)$ ,  $\mathcal{I}_3 = [3/2, 2 - 4\epsilon/\sqrt{3})$ ,  $\mathcal{I}_4 = [2 - 4\epsilon/\sqrt{3}, 2)$ ,  $\mathcal{I}_5 = [2, \sqrt{3}/(2\epsilon) - 1)$ ,  $\mathcal{I}_6 = [\sqrt{3}/(2\epsilon) - 1, \sqrt{3}/(2\epsilon))$ , and  $\mathcal{I}_7 = [\sqrt{3}/(2\epsilon), \infty)$ .

For  $\epsilon \in [\sqrt{3}/8, \sqrt{3}/6)$ ,  $\mu_S(r, \epsilon) = \sum_{j=1}^7 \varpi_{2,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$  where  $\varpi_{2,j}(r, \epsilon) = \varpi_{1,j}(r, \epsilon)$  for  $j = 1, 2, 4, 5, 6$ , and for  $j = 3, 7$ ,

$$\begin{aligned}
\varpi_{2,3}(r, \epsilon) &= - \left[ 576 r^4 \epsilon^4 - 1152 r^2 \epsilon^4 + 37 r^4 + 224 \sqrt{3} r^3 \epsilon + 864 r^2 \epsilon^2 - 384 \epsilon^4 - 336 r^3 - 576 \sqrt{3} r^2 \epsilon \right. \\
&\quad \left. + 768 \sqrt{3} \epsilon^3 + 432 r^2 - 1728 \epsilon^2 + 576 \sqrt{3} \epsilon - 216 \right] / \left[ 216 r^2 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2 \right], \\
\varpi_{2,7}(r, \epsilon) &= 1,
\end{aligned}$$

with the corresponding intervals  $\mathcal{I}_1 = [1, 3/2 - \sqrt{3}\epsilon)$ ,  $\mathcal{I}_2 = [3/2 - \sqrt{3}\epsilon, 2 - 4\epsilon/\sqrt{3})$ ,  $\mathcal{I}_3 = [2 - 4\epsilon/\sqrt{3}, 3/2)$ ,  $\mathcal{I}_4 = [3/2, 2)$ ,  $\mathcal{I}_5 = [2, \sqrt{3}/(2\epsilon) - 1)$ ,  $\mathcal{I}_6 = [\sqrt{3}/(2\epsilon) - 1, \sqrt{3}/(2\epsilon))$ , and  $\mathcal{I}_5 = [\sqrt{3}/(2\epsilon), \infty)$ .

For  $\epsilon \in [\sqrt{3}/6, \sqrt{3}/4)$ ,  $\mu_S(r, \epsilon) = \sum_{j=1}^6 \varpi_{3,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$  where  $\varpi_{3,1}(r, \epsilon) = \varpi_{1,2}(r, \epsilon)$  and

$$\begin{aligned}
\varpi_{3,2}(r, \epsilon) &= - \left[ 576 r^4 \epsilon^4 - 1152 r^2 \epsilon^4 + 37 r^4 + 224 \sqrt{3} r^3 \epsilon + 864 r^2 \epsilon^2 - 384 \epsilon^4 - 336 r^3 - 576 \sqrt{3} r^2 \epsilon \right. \\
&\quad \left. + 768 \sqrt{3} \epsilon^3 + 432 r^2 - 1728 \epsilon^2 + 576 \sqrt{3} \epsilon - 216 \right] / \left[ 216 r^2 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2 \right], \\
\varpi_{3,3}(r, \epsilon) &= \left[ 576 r^2 \epsilon^4 + 3072 r \epsilon^4 - 1536 \sqrt{3} r \epsilon^3 + 3456 \epsilon^4 - 2304 \sqrt{3} \epsilon^3 - 37 r^2 - 224 \sqrt{3} r \epsilon \right. \\
&\quad \left. + 864 \epsilon^2 + 336 r + 576 \sqrt{3} \epsilon - 432 \right] / \left[ 216 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2 \right],
\end{aligned}$$

$$\begin{aligned}
\varpi_{3,4}(r, \epsilon) &= \left[ 192 r^4 \epsilon^4 + 1024 r^3 \epsilon^4 - 512 \sqrt{3} r^3 \epsilon^3 + 1152 r^2 \epsilon^4 - 768 \sqrt{3} r^2 \epsilon^3 + 9 r^4 + 96 \sqrt{3} r^3 \epsilon + 288 r^2 \epsilon^2 \right. \\
&\quad \left. - 144 r^3 - 576 \sqrt{3} r^2 \epsilon + 720 r^2 + 1152 \sqrt{3} r \epsilon - 1152 r - 576 \sqrt{3} \epsilon + 540 \right] / \left[ 72 r^2 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2 \right], \\
\varpi_{3,5}(r, \epsilon) &= \frac{48 r^4 \epsilon^4 + 256 r^3 \epsilon^4 - 128 \sqrt{3} r^3 \epsilon^3 + 288 r^2 \epsilon^4 - 192 \sqrt{3} r^2 \epsilon^3 + 72 r^2 \epsilon^2 + 18 r^2 + 48 \sqrt{3} \epsilon - 45}{18 r^2 (2 \epsilon + 1)^2 (2 \epsilon - 1)^2}, \\
\varpi_{3,6}(r, \epsilon) &= 1,
\end{aligned}$$

with the corresponding intervals  $\mathcal{I}_1 = [1, 2 - 4\epsilon/\sqrt{3})$ ,  $\mathcal{I}_2 = [2 - 4\epsilon/\sqrt{3}, \sqrt{3}/(2\epsilon) - 1)$ ,  $\mathcal{I}_3 = [\sqrt{3}/(2\epsilon) - 1, 3/2)$ ,  $\mathcal{I}_4 = [3/2, 2)$ ,  $\mathcal{I}_5 = [2, \sqrt{3}/(2\epsilon))$ , and  $\mathcal{I}_5 = [\sqrt{3}/(2\epsilon), \infty)$ .

For  $\epsilon \in [\sqrt{3}/4, \sqrt{3}/3)$ ,  $\mu_S(r, \epsilon) = \sum_{j=1}^3 \varpi_{4,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$  where

$$\begin{aligned}\varpi_{4,1}(r, \epsilon) &= -\frac{9r^2\epsilon^2 + 2\sqrt{3}r^2\epsilon + 48r\epsilon^2 + r^2 - 16\sqrt{3}r\epsilon - 90\epsilon^2 - 12r + 36\sqrt{3}\epsilon}{18(3\epsilon - \sqrt{3})^2}, \\ \varpi_{4,2}(r, \epsilon) &= -\left[9r^4\epsilon^4 - 4\sqrt{3}r^4\epsilon^3 + 48r^3\epsilon^4 - 48\sqrt{3}r^3\epsilon^3 - 90r^2\epsilon^4 + 36r^3\epsilon^2 + 96\sqrt{3}r^2\epsilon^3 - 126r^2\epsilon^2 \right. \\ &\quad \left. - 32\sqrt{3}r\epsilon^3 - 48\epsilon^4 + 36\sqrt{3}r^2\epsilon + 144r\epsilon^2 + 96\sqrt{3}\epsilon^3 - 18r^2 - 72\sqrt{3}r\epsilon - 216\epsilon^2 + 36r \right. \\ &\quad \left. + 72\sqrt{3}\epsilon - 27\right] / \left[2(3\epsilon - \sqrt{3})^4 r^2\right], \\ \varpi_{4,3}(r, \epsilon) &= 1,\end{aligned}$$

with the corresponding intervals  $\mathcal{I}_1 = [1, 3 - 2\epsilon/\sqrt{3})$ ,  $\mathcal{I}_2 = [3 - 2\epsilon/\sqrt{3}, \sqrt{3}/\epsilon - 2)$ , and  $\mathcal{I}_3 = [\sqrt{3}/\epsilon - 2, \infty)$ .

### $\mu_A(r, \epsilon)$ Under Association Alternatives

Under association, we compute  $\mu_A(r, \epsilon)$  explicitly. For  $\epsilon \in [0, (7\sqrt{3} - 3\sqrt{15})/12 \approx .042)$ ,  $\mu_A(r, \epsilon) = \sum_{j=1}^6 \varpi_{1,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$  where

$$\begin{aligned}\varpi_{1,1}(r, \epsilon) &= -\left[3456\epsilon^4 r^4 + 9216\epsilon^4 r^3 - 3072\sqrt{3}\epsilon^3 r^4 - 17280\epsilon^4 r^2 - 3072\sqrt{3}\epsilon^3 r^3 + 2304\epsilon^2 r^4 \right. \\ &\quad \left. + 4608\sqrt{3}\epsilon^3 r^2 - 2304\epsilon^2 r^3 + 6336\epsilon^4 + 6144\sqrt{3}\epsilon^3 r + 6912\epsilon^2 r^2 + 512\sqrt{3}\epsilon r^3 \right. \\ &\quad \left. - 101r^4 - 6144\sqrt{3}\epsilon^3 - 11520\epsilon^2 r - 1536\sqrt{3}\epsilon r^2 + 256r^3 + 5760\epsilon^2 + 1536\sqrt{3}\epsilon r \right. \\ &\quad \left. - 384r^2 - 512\sqrt{3}\epsilon + 256r - 64\right] / \left[24(6\epsilon + \sqrt{3})^2(6\epsilon - \sqrt{3})^2 r^2\right], \\ \varpi_{1,2}(r, \epsilon) &= -\left[1728\epsilon^4 r^4 - 1536\sqrt{3}\epsilon^3 r^4 - 31104\epsilon^4 r^2 + 1152\epsilon^2 r^4 + 15552\epsilon^4 + 10368\epsilon^2 r^2 - 37r^4 \right. \\ &\quad \left. - 20736\epsilon^2 r + 10368\epsilon^2\right] / \left[24(6\epsilon + \sqrt{3})^2(6\epsilon - \sqrt{3})^2 r^2\right], \\ \varpi_{1,3}(r, \epsilon) &= \left[-2592\epsilon^4 r^4 - 2304\sqrt{3}\epsilon^3 r^4 - 46656\epsilon^4 r^2 + 1728\epsilon^2 r^4 + 10656\epsilon^4 - 9216\sqrt{3}\epsilon^3 r \right. \\ &\quad \left. + 9072\epsilon^2 r^2 - 432\sqrt{3}\epsilon r^3 - 15r^4 + 12288\sqrt{3}\epsilon^3 - 13824\epsilon^2 r + 1728\sqrt{3}\epsilon r^2 - 216r^3 \right. \\ &\quad \left. + 4032\epsilon^2 - 2304\sqrt{3}\epsilon r + 432r^2 + 1024\sqrt{3}\epsilon - 384r + 128\right] / \left[36(6\epsilon + \sqrt{3})^2(6\epsilon - \sqrt{3})^2 r^2\right], \\ \varpi_{1,4}(r, \epsilon) &= -\frac{1728\epsilon^4 r^4 - 1536\sqrt{3}\epsilon^3 r^4 - 31104\epsilon^4 r^2 + 1152\epsilon^2 r^4 - 5184\epsilon^4 + 2592\epsilon^2 r^2 - 37r^4 - 3456\epsilon^2}{24(6\epsilon + \sqrt{3})^2(6\epsilon - \sqrt{3})^2 r^2}, \\ \varpi_{1,5}(r, \epsilon) &= \frac{9}{8} \frac{1152\epsilon^4 r^2 + 192\epsilon^4 - 192\epsilon^2 r^2 - r^4 + 128\epsilon^2 + 32r^2 - 64r + 36}{(6\epsilon + \sqrt{3})^2(6\epsilon - \sqrt{3})^2 r^2}, \\ \varpi_{1,6}(r, \epsilon) &= -\frac{9}{8} \frac{(r+6)(r-2)^3}{(6\epsilon + \sqrt{3})^2(6\epsilon - \sqrt{3})^2 r^2},\end{aligned}$$

with the corresponding intervals  $\mathcal{I}_1 = \left[1, \frac{1+2\sqrt{3}\epsilon}{1-\sqrt{3}\epsilon}\right)$ ,  $\mathcal{I}_2 = \left[\frac{1+2\sqrt{3}\epsilon}{1-\sqrt{3}\epsilon}, \frac{4(1-\sqrt{3}\epsilon)}{3}\right)$ ,  $\mathcal{I}_3 = \left[\frac{4(1-\sqrt{3}\epsilon)}{3}, \frac{4(1+2\sqrt{3}\epsilon)}{3}\right)$ ,  $\mathcal{I}_4 = \left[\frac{4(1+2\sqrt{3}\epsilon)}{3}, \frac{3}{2(1-\sqrt{3}\epsilon)}\right)$ ,  $\mathcal{I}_5 = \left[\frac{3}{2(1-\sqrt{3}\epsilon)}, 2\right)$  and  $\mathcal{I}_6 = [2, \infty)$ .

For  $\epsilon \in \left[(7\sqrt{3} - 3\sqrt{15})/12, \sqrt{3}/12\right)$ ,  $\mu_A(r, \epsilon) = \sum_{j=1}^6 \varpi_{2,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$  where  $\varpi_{2,j}(r, \epsilon) = \varpi_{1,j}(r, \epsilon)$  for  $j = 1, 3, 4, 5, 6$  and

$$\begin{aligned}\varpi_{2,2}(r, \epsilon) &= \left[-3456\epsilon^2 r^4 + 111r^4 - 5184\epsilon^4 r^4 + 4608\sqrt{3}\epsilon^3 r^4 - 336\sqrt{3}\epsilon r^3 - 168r^3 - 13824\epsilon^4 r^3 \right. \\ &\quad \left. + 4608\sqrt{3}\epsilon^3 r^3 + 3456\epsilon^2 r^3 + 144r^2 - 6912\sqrt{3}\epsilon^3 r^2 - 3888\epsilon^2 r^2 + 576\sqrt{3}\epsilon r^2 + 25920\epsilon^4 r^2 \right. \\ &\quad \left. + 3168\epsilon^4 + 2880\epsilon^2 - 256\sqrt{3}\epsilon - 32 - 3072\sqrt{3}\epsilon^3\right] / \left[36(\sqrt{3} + 6\epsilon)^2(-6\epsilon + \sqrt{3})^2 r^2\right]\end{aligned}$$

with the corresponding intervals  $\mathcal{I}_1 = \left[1, \frac{4(1-\sqrt{3}\epsilon)}{3}\right)$ ,  $\mathcal{I}_2 = \left[\frac{4(1-\sqrt{3}\epsilon)}{3}, \frac{1+2\sqrt{3}\epsilon}{1-\sqrt{3}\epsilon}\right)$ ,  $\mathcal{I}_3 = \left[\frac{1+2\sqrt{3}\epsilon}{1-\sqrt{3}\epsilon}, \frac{4(1+2\sqrt{3}\epsilon)}{3}\right)$ ,  $\mathcal{I}_4 = \left[\frac{4(1+2\sqrt{3}\epsilon)}{3}, \frac{3}{2(1-\sqrt{3}\epsilon)}\right)$ ,  $\mathcal{I}_5 = \left[\frac{3}{2(1-\sqrt{3}\epsilon)}, 2\right)$  and  $\mathcal{I}_6 = [2, \infty)$ .

For  $\epsilon \in \left[\sqrt{3}/12, \sqrt{3}/3\right)$ ,  $\mu_A(r, \epsilon) = \sum_{j=1}^3 \varpi_{3,j}(r, \epsilon) \mathbf{I}(r \in \mathcal{I}_j)$  where

$$\begin{aligned}\varpi_{3,1}(r, \epsilon) &= \frac{2r^2 - 1}{6r^2}, \\ \varpi_{3,2}(r, \epsilon) &= \left[432\epsilon^4 r^4 + 1152\epsilon^4 r^3 - 576\sqrt{3}\epsilon^3 r^4 + 1296\epsilon^4 r^2 - 960\sqrt{3}\epsilon^3 r^3 + 864\epsilon^2 r^4 - 864\sqrt{3}\epsilon^3 r^2 \right. \\ &\quad \left. + 576\epsilon^2 r^3 - 192\sqrt{3}\epsilon r^4 - 360\epsilon^4 + 648\epsilon^2 r^2 + 64\sqrt{3}\epsilon r^3 + 48r^4 + 192\sqrt{3}\epsilon^3 - 144\sqrt{3}\epsilon r^2 \right. \\ &\quad \left. - 64r^3 - 504\epsilon^2 + 72r^2 + 88\sqrt{3}\epsilon - 25\right] / \left[16(3\epsilon - \sqrt{3})^4 r^2\right], \\ \varpi_{3,3}(r, \epsilon) &= -\frac{-54\epsilon^2 r^2 + 36\sqrt{3}\epsilon r^2 + 15\epsilon^2 - 18r^2 + 2\sqrt{3}\epsilon + 20}{6(-3\epsilon + \sqrt{3})^2 r^2},\end{aligned}$$

with the corresponding intervals  $\mathcal{I}_1 = \left[1, \frac{1+2\sqrt{3}\epsilon}{2(1-\sqrt{3}\epsilon)}\right)$ ,  $\mathcal{I}_3 = \left[\frac{1+2\sqrt{3}\epsilon}{2(1-\sqrt{3}\epsilon)}, \frac{3}{2(1-\sqrt{3}\epsilon)}\right)$ ,  $\mathcal{I}_5 = \left[\frac{3}{2(1-\sqrt{3}\epsilon)}, \infty\right)$ .

### Appendix 3: Derivation of $\mu(r, \epsilon)$

We demonstrate the derivation of  $\mu_S(r, \epsilon)$  for segregation with  $\epsilon \in [0, \sqrt{3}/8)$  and among the intervals of  $r$  that do not vanish as  $\epsilon \rightarrow 0$ . So the resultant expressions can be used in PAE analysis.

First, observe that, by symmetry,

$$\mu_S(r, \epsilon) = P(X_2 \in N_Y^r(X_1, \epsilon)) = 6P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in T_s \setminus T(y_1, \epsilon)).$$

Let  $q(y_j, x)$  be the line parallel to  $e_j$  and crossing  $T(\mathcal{Y})$  such that  $d(y_j, q(y_j, x)) = \epsilon$  for  $j = 1, 2, 3$ . Furthermore, let  $T_\epsilon := T(\mathcal{Y}) \setminus \cup_{j=1}^3 T(y_j, \epsilon)$ . Then  $q(y_1, x) = 2\epsilon - \sqrt{3}x$ ,  $q(y_2, x) = \sqrt{3}x - \sqrt{3} + 2\epsilon$ , and  $q(y_3, x) = \sqrt{3}/2 - \epsilon$ . Now, let

$$\begin{aligned}Q_1 &= q(y_1, x) \cap \overline{y_1 y_2} = (2\epsilon/\sqrt{3}, 0), & Q_2 &= q(y_2, x) \cap \overline{y_1 y_2} = (1 - 2\epsilon/\sqrt{3}, 0), \\ Q_3 &= q(y_2, x) \cap \overline{y_2 y_3} = (1 - \epsilon/\sqrt{3}, \epsilon), & Q_4 &= q(y_3, x) \cap \overline{y_2 y_3} = (1/2 + \epsilon/\sqrt{3}, \sqrt{3}/2 - \epsilon), \\ Q_5 &= q(y_3, x) \cap \overline{y_1 y_3} = (1/2 - \epsilon/\sqrt{3}, \sqrt{3}/2 - \epsilon), & Q_6 &= q(y_1, x) \cap \overline{y_1 y_3} = (\epsilon/\sqrt{3}, \epsilon).\end{aligned}$$

See Figure 43. Then  $T(y_1, \epsilon) = T(y_1, Q_1, Q_6)$ ,  $T(y_2, \epsilon) = T(Q_2, y_2, Q_3)$ , and  $T(y_3, \epsilon) = T(Q_4, Q_5, y_3)$ , and for  $\epsilon \in [0, \sqrt{3}/4)$ ,  $T_\epsilon$  is the hexagon with vertices,  $Q_j$ ,  $j = 1, \dots, 6$ . Now, let  $q_2(x)$  be the line such that  $rd(y_1, q_2(x)) = d(y_1, \ell(Q_2)) = d(y_1, \overline{y_2 y_3}) - \epsilon$ ,  $q_3(x)$  be the line such that  $rd(y_1, q_3(x)) = d(y_1, \overline{y_2 y_3})$ . Then  $q_2(x) = -\sqrt{3}x + (\sqrt{3} - 2\epsilon)/r$  and  $q_3(x)$  is the same as  $\ell_s(x)$  before. Let the  $x$  coordinate of  $q(y_1, x) \cap \ell_{am}(x)$  be  $s_1$ ,  $q_2(x) \cap \ell_{am}(x)$  be  $s_3$ , and  $\ell_s(x) \cap \ell_{am}(x)$  be  $s_5$  and  $Q_1 = (s_2, 0)$ ,  $q_2(x) \cap \overline{y_1 y_2} = (s_4, 0)$ , and  $\ell_s(x) \cap \overline{y_1 y_2} = (s_6, 0)$ . So  $s_1 := \sqrt{3}\epsilon/2$ ,  $s_2 = 2\epsilon/\sqrt{3}$ ,  $s_3 = (3 - 2\epsilon\sqrt{3})/(4r)$ ,  $s_4 = (3 - 2\epsilon\sqrt{3})/(3r)$ ,  $s_5 = 3/(4r)$ , and  $s_6 = 1/r$ .

See Figure 44 for an  $r \in [2, \sqrt{3}/(2\epsilon)]$ . Furthermore, for  $x_1 = (x, y) \in R_{CM}(y_1)$ , let

$$\begin{aligned}U_1 &:= q(y_2, x) \cap \ell_r(x_1, x) = \left(\left(\sqrt{3}y/6 + x/2\right)r + 1/2 - \epsilon/\sqrt{3}, \left(y/2 + \sqrt{3}x/2\right)r - \sqrt{3}/2 + \epsilon\right), \text{ and} \\ U_2 &:= q(y_3, x) \cap \ell_r(x_1, x) = \left(\left(y/\sqrt{3} + x\right)r + \epsilon/\sqrt{3} - 1/2, \sqrt{3}/2 - \epsilon\right).\end{aligned}$$

Let  $\mathcal{P}(a_1, a_2, \dots, a_n)$  denote the polygon with vertices  $a_1, a_2, \dots, a_n$ . If  $x_1$  is below  $q_2(x)$ , then  $N_Y^r(x_1, \epsilon) = A(N_Y^r(x_1)) \setminus T(y_1, \epsilon) = \mathcal{P}(Q_1, N_1, N_2, Q_6)$ , if  $x_1$  is between  $q_2(x)$  and  $\ell_s(x)$ , then  $N_Y^r(x_1, \epsilon) = \mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)$ , and if  $x_1$  is above  $\ell_s(x)$ , then  $N_Y^r(x_1, \epsilon) = T_\epsilon$ .

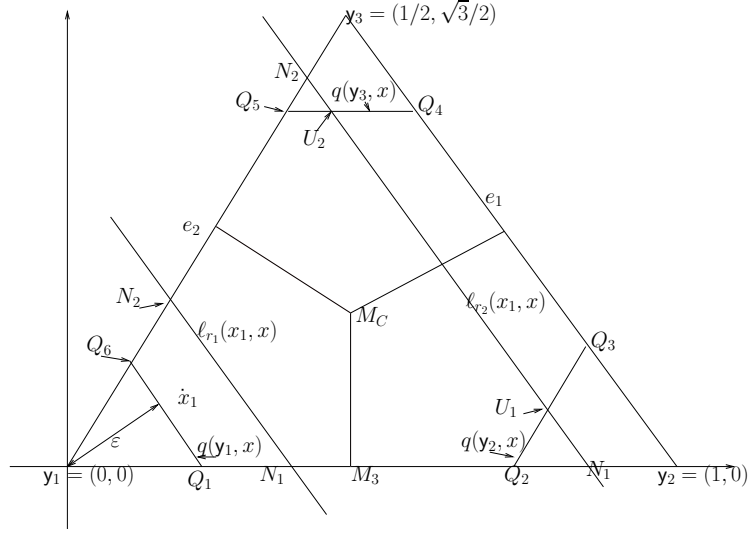


Figure 43: The support under  $H_\epsilon^S$  for  $\epsilon \in (0, \sqrt{3}/4)$  and the two types  $\ell_r(x_1, x)$  for  $r_1 < r_2$ .

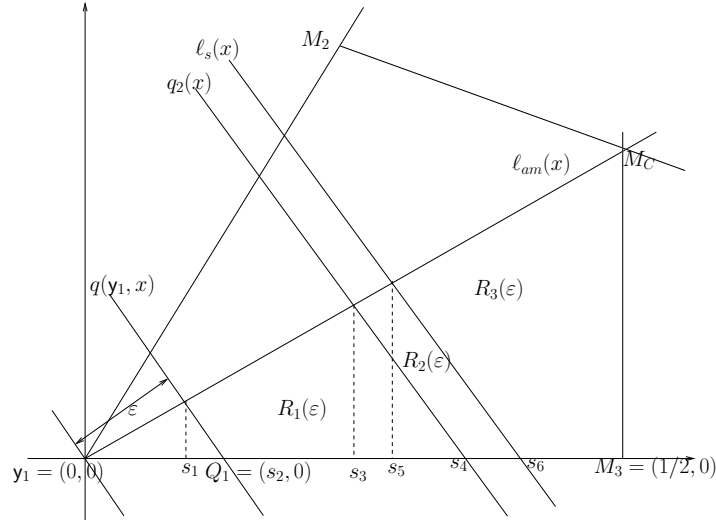


Figure 44: The partition of  $T_s$  for different types of  $N_y^r(\cdot, \epsilon)$  under  $H_\epsilon^S$  with  $r \in [2, \sqrt{3}/(2\epsilon))$ .

For  $r \in [1, 3/2 - \sqrt{3}\epsilon)$ , since  $\epsilon$  small enough that  $q_2(x) \cap T_s = \emptyset$ , then  $N(x, \epsilon) \subsetneq T_\epsilon$  for all  $x \in T_s \setminus T(y_1, \epsilon)$ . Then

$$\begin{aligned} P(X_2 \in N_y^r(X_1, \epsilon), X_1 \in T_s \setminus T(y_1, \epsilon)) &= \int_{s_1}^{s_2} \int_{q(y_1, x)}^{\ell_{am}(x)} \frac{A(N_y^r(x_1, \epsilon))}{A(T_\epsilon)^2} dy dx \\ &+ \int_{s_2}^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_y^r(x_1, \epsilon))}{A(T_\epsilon)^2} dy dx = -\frac{(576r^2 - 1152)\epsilon^4 + 288\epsilon^2 - 37r^2}{1296(2\epsilon - 1)^2(2\epsilon + 1)^2}. \end{aligned}$$

where  $A(N_y^r(x_1, \epsilon)) = A(\mathcal{P}(Q_1, N_1, N_2, Q_6)) = \left(\frac{\sqrt{3}}{12}y^2 + \frac{1}{2}xy + \frac{\sqrt{3}}{4}x^2\right)r^2 - \frac{\sqrt{3}}{3}\epsilon^2$  and  $A(T_\epsilon) = \sqrt{3}/4 - \sqrt{3}\epsilon^2$  and  $\ell_{am}(x) = x/\sqrt{3}$  is the equation of the line segment  $\overline{y_1 M_C}$ . Hence for  $r \in [1, 3/2)$ ,  $\mu(r) = -\frac{(576r^2 - 1152)\epsilon^4 + 288\epsilon^2 - 37r^2}{216(2\epsilon - 1)^2(2\epsilon + 1)^2}$ .

For  $r \in [3/2, 2 - 4\epsilon/\sqrt{3})$ ,  $\ell_s(x)$  crosses through  $\overline{M_3 M_C}$ . Since  $\epsilon$  small enough so that  $q_2(x)$  does the same. So  $x_1$  below  $q_2(x)$  is equivalent to  $x_1 \in R_1(\epsilon)$  where

$$R_1(\epsilon) = \{(x, y) \in [s_3, s_5] \times [q_2(x), \ell_{am}(x)] \cup [s_2, s_3] \times [q_2(x), \ell_{am}(x)] \cup [s_3, 1/2] \times [0, q_2(x)]\}.$$

Then

$$\begin{aligned}
P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_1(\epsilon)) &= \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_Y^r(x_1, \epsilon))}{A(T_\epsilon)^2} dy dx \\
&= \int_{s_1}^{s_2} \int_{q(y_1, x)}^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx + \int_{s_2}^{s_3} \int_0^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx \\
&\quad + \int_{s_3}^{1/2} \int_0^{q_2(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx = \left[ (384r^2 + 576 - 192r^4)\epsilon^4 + 512\epsilon^3\sqrt{3}r + \right. \\
&\quad \left. (288r^2 - 1728)\epsilon^2 + (-576\sqrt{3}r + 864\sqrt{3})\epsilon - 9r^4 - 324 + 288r \right] / \left[ 432((2\epsilon - 1)^2(2\epsilon + 1)^2r^2) \right]
\end{aligned}$$

where  $A(\mathcal{P}(Q_1, N_1, N_2, Q_6))$  is same as before.

Next,  $x_1$  between  $q_2(x)$  and  $\ell_s(x)$  is equivalent to  $x_1 \in R_2(\epsilon)$  where

$$R_2(\epsilon) = \{(x, y) \in [s_9, s_2] \times [q_1(x), \ell_{am}(x)] \cup [s_5, 1/2] \times [q_2(x), \ell_s(x)]\}.$$

Then

$$\begin{aligned}
P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_2(\epsilon)) &= \int_{s_3}^{s_5} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))}{A(T_\epsilon)^2} dy dx \\
&\quad + \int_{s_5}^{1/2} \int_0^{q_2(x)} \frac{A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))}{A(T_\epsilon)^2} dy dx = -\frac{2\sqrt{3}\epsilon(10\epsilon^3\sqrt{3} + (32r - 24)\epsilon^2 + (-27\sqrt{3} + 12\sqrt{3}r)\epsilon - 18r + 27)}{27(4\epsilon^2 - 1)^2r^2}
\end{aligned}$$

where

$$\begin{aligned}
A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)) &= -\sqrt{3}\epsilon^2 + (2 - 2ry/\sqrt{3} - 2rx)\epsilon + ry + \sqrt{3}rx - \sqrt{3}/2 - \sqrt{3}r^2y^2/12 \\
&\quad - r^2xy/2 - \sqrt{3}r^2x^2/4.
\end{aligned}$$

Furthermore,  $x_1$  above  $\ell_s(x)$  is equivalent to  $x_1 \in R_3(\epsilon)$  where  $R_3(\epsilon) = \{(x, y) \in [s_5, 1/2] \times [\ell_{am}(x), \ell_s(x)]\}$ .

$$\text{Then } P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_3(\epsilon)) = \int_{s_5}^{1/2} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T_\epsilon)} dy dx = -\frac{(2r-3)^2}{6(2\epsilon-1)(2\epsilon+1)r^2}.$$

Hence for  $r \in [3/2, 2)$ ,

$$\begin{aligned}
\mu_S(r, \epsilon) &= 6 \left( - \left[ (-384r^2 + 384 + 192r^4)\epsilon^4 + (-768\sqrt{3} + 512\sqrt{3}r)\epsilon^3 + (1728 - 2304r + 864r^2)\epsilon^2 \right. \right. \\
&\quad \left. \left. - 288r^2 - 324 + 9r^4 + 576r \right] / \left[ 432((2\epsilon - 1)^2(2\epsilon + 1)^2r^2) \right] \right) \\
&= - \left[ (-384r^2 + 384 + 192r^4)\epsilon^4 + (-768\sqrt{3} + 512\sqrt{3}r)\epsilon^3 + (1728 - 2304r + 864r^2)\epsilon^2 \right. \\
&\quad \left. - 288r^2 - 324 + 9r^4 + 576r \right] / \left[ 72((2\epsilon + 1)^2(2\epsilon - 1)^2r^2) \right].
\end{aligned}$$

For  $r \in [2, \infty)$ ,  $\ell_s(x)$  crosses through  $\overline{y_1M_3}$ , so the same types of  $N_Y^r(x_1, \epsilon)$  occur as above. The explicit forms of  $R_j(\epsilon)$ ,  $j = 1, 2, 3$  change and are given below:

$$\begin{aligned}
R_1(\epsilon) &= \{(x, y) \in [s_1, s_2] \times [q(y_1, x), \ell_{am}(x)] \cup [s_2, s_3] \times [0, \ell_{am}(x)] \cup [s_3, s_4] \times [0, q_2(x)]\} \\
R_2(\epsilon) &= \{(x, y) \in [s_3, s_5] \times [q_2(x), \ell_{am}(x)] \cup [s_5, s_4] \times [q_2(x), \ell_s(x)] \cup [s_4, s_6] \times [0, \ell_s(x)]\} \\
R_3(\epsilon) &= \{(x, y) \in [s_5, s_6] \times [\ell_s(x), \ell_{am}(x)] \cup [s_6, 1/2] \times [0, \ell_{am}(x)]\}.
\end{aligned}$$

Then

$$\begin{aligned}
P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_1(\epsilon)) &= \int_{s_1}^{s_2} \int_{q(y_1, x)}^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx \\
&\quad + \int_{s_2}^{s_3} \int_0^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx + \int_{s_3}^{s_4} \int_0^{q_2(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx \\
&= -\frac{9 + (-32r^2 + 16r^4 + 16)\epsilon^4 - 48\epsilon^2 + 24\epsilon\sqrt{3}}{36(4\epsilon^2 - 1)^2r^2}.
\end{aligned}$$

where  $A(\mathcal{P}(Q_1, N_1, N_2, Q_6))$  is same as before.

Furthermore,

$$\begin{aligned} P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_2(\epsilon)) &= \int_{s_3}^{s_5} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))}{A(T_\epsilon)^2} dy dx \\ &\quad + \int_{s_5}^{s_4} \int_{q_2(x)}^{\ell_s(x)} \frac{A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))}{A(T_\epsilon)^2} dy dx + \int_{s_4}^{s_6} \int_0^{\ell_s(x)} \frac{A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))}{A(T_\epsilon)^2} dy dx \\ &= \frac{2\sqrt{3}\epsilon(-27\epsilon\sqrt{3} - 24\epsilon^2 + 10\epsilon^3\sqrt{3} + 27)}{81(4\epsilon^2 - 1)^2 r^2}. \end{aligned}$$

where  $A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))$  is same as before.

Next,

$$P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_3(\epsilon)) = \int_{s_5}^{s_6} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T_\epsilon)} dy dx + \int_{s_6}^{1/2} \int_0^{\ell_{am}(x)} \frac{1}{A(T_\epsilon)} dy dx = \frac{3-r^2}{6(4\epsilon^2-1)r^2}.$$

Hence for  $r \in [2, \sqrt{3}/(2\epsilon) - 1]$ ,

$$\begin{aligned} \mu_S(r, \epsilon) &= 6 \left( -\frac{(48r^4 - 32 - 96r^2)\epsilon^4 + 64\epsilon^3\sqrt{3} + (72r^2 - 144)\epsilon^2 + 27 - 18r^2}{108(4\epsilon^2 - 1)^2 r^2} \right) \\ &= -\frac{(48r^4 - 32 - 96r^2)\epsilon^4 + 64\epsilon^3\sqrt{3} + (72r^2 - 144)\epsilon^2 + 27 - 18r^2}{18(4\epsilon^2 - 1)^2 r^2}. \end{aligned}$$

For  $r = \infty$ , it is trivial to see that  $\mu(r) = 1$ . In fact, for fixed  $\epsilon > 0$ ,  $\mu(r) = 1$  for  $r \geq \sqrt{3}/(2\epsilon)$ .

## Appendix 4: Derivation of $\mu_S(r, \epsilon)$ and $\nu_S(r, \epsilon)$ for Segregation with $\epsilon = \sqrt{3}/8$

For the segregation alternative with  $\epsilon = \sqrt{3}/8$ ,  $\mu_S(r, \epsilon = \sqrt{3}/8) = 1$  for  $r \geq 4$ , so we find  $\mu_S(r, \epsilon = \sqrt{3}/8)$  for  $r \in [1, 4)$ . In particular, for the mean we partition  $[1, 4)$  into five intervals,  $[1, 9/8)$ ,  $[9/8, 3/2)$ ,  $[3/2, 2)$ ,  $[2, 3)$ ,  $[3, 4)$ , and for the covariance into twelve intervals,  $[1, 12/11)$ ,  $[12/11, 9/8)$ ,  $[9/8, \sqrt{6}/11)$ ,  $[\sqrt{6}/11, 21/16)$ ,  $[21/16, 4/3)$ ,  $[4/3, 3/2)$ ,  $[3/2, \sqrt{3})$ ,  $[\sqrt{3}, 7/4)$ ,  $[7/4, 2)$ ,  $[2, 3)$ ,  $[3, 7/2)$ ,  $[7/2, 4)$ . We pick the sample intervals  $[3/2, 2)$  and  $[7/4, 2)$  to demonstrate the calculations of the mean and the variance, respectively. Then observe that, by symmetry,

$$\mu_S(r, \epsilon) = P(X_2 \in N_Y^r(X_1, \epsilon)) = 6 P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_\epsilon(y_1)).$$

Then  $q(y_1, x) = \sqrt{3}(1/4 - x)$ ,  $q(y_2, x) = \sqrt{3}(4x - 3)/4$ , and  $q(y_3, x) = 3\sqrt{3}/8$ . See Section ?? for the definition of  $q(y_j, x)$ . Hence,  $Q_1 = q(y_1, x) \cap \overline{y_1 y_2} = (1/4, 0)$ ,  $Q_2 = q(y_2, x) \cap \overline{y_1 y_2} = (3/4, 0)$ ,  $Q_3 = q(y_2, x) \cap \overline{y_2 y_3} = (7/8, 3/8)$ ,  $Q_4 = q(y_3, x) \cap \overline{y_2 y_3} = (5/8, 3\sqrt{3}/8)$ ,  $Q_5 = q(y_3, x) \cap \overline{y_1 y_3} = (3/8, 3\sqrt{3}/8)$ ,  $Q_6 = q(y_1, x) \cap \overline{y_1 y_3} = (1/8, 3/8)$ . Then  $T(y_1, \epsilon) = T(y_1, Q_1, Q_6)$ ,  $T(y_2, \epsilon) = T(Q_2, y_2, Q_3)$ , and  $T(y_3, \epsilon) = T(Q_4, Q_5, y_3)$ , and for  $\epsilon = \sqrt{3}/8$ ,  $T_\epsilon$  is the hexagon with vertices,  $Q_j$ ,  $j = 1, \dots, 6$ .

Furthermore,  $q_2(x) = -\sqrt{3}(x - 3/(4r))$  and  $q_3(x) = \sqrt{3}(1/r - x)$  (i.e. the same as  $\ell_s(x)$  before). See Section ?? for the definition of  $q_j(x)$ . Let the  $x$  coordinate of  $q(y_1, x) \cap \ell_{am}(x)$  be  $s_1$ ,  $q_2(x) \cap \ell_{am}(x)$  be  $s_4$ , and  $\ell_s(x) \cap \ell_{am}(x)$  be  $s_{10}$  and  $Q_1 = (s_3, 0)$ ,  $q_2(x) \cap \overline{y_1 y_2} = (s_6, 0)$ , and  $\ell_s(x) \cap \overline{y_1 y_2} = (s_{12}, 0)$ . So  $s_1 = 3/16$  and  $s_3 = 1/4$ ,  $s_4 := 9/(16r)$ ,  $s_6 = 3/(4r) = s_{10}$ , and  $s_{12} = 1/r$ . See Figure 45.

For  $x_1 = (x, y) \in R_{CM}(y_1)$ , let

$$\begin{aligned} U_1 &:= q(y_2, x) \cap \ell_r(x_1, x) = \left( \sqrt{3} \left( 4ry + 4\sqrt{3}rx + 3\sqrt{3} \right) / 24, ry/2 + \sqrt{3}rx/2 - 3\sqrt{3}/8 \right), \text{ and} \\ U_2 &:= q(y_3, x) \cap \ell_r(x_1, x) = \left( \sqrt{3} \left( 8ry + 8\sqrt{3}rx - 3\sqrt{3} \right) / 24, 3\sqrt{3}/8 \right). \end{aligned}$$

If  $x_1$  is below  $q_2(x)$ , then  $N_Y^r(x_1, \epsilon) = N_Y^r(x_1) \cap T(y_1, \sqrt{3}/8) = \mathcal{P}(Q_1, N_1, N_2, Q_6)$ , if  $x_1$  is between  $q_2(x)$  and  $\ell_s(x)$ , then  $N_Y^r(x_1, \epsilon) = \mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)$ , and if  $x_1$  is above  $\ell_s(x)$ , then  $N_Y^r(x_1, \epsilon) = T_\epsilon$ .

For  $r \in [3/2, 2)$ ,  $q_3(x)$  crosses through  $\overline{M_3 M_C}$  and  $q_2(x)$  crosses  $\overline{y_1 M_3}$ . So  $x_1$  below  $q_2(x)$  is equivalent to  $x_1 \in R_1(\sqrt{3}/8)$  where

$$R_1(\sqrt{3}/8) = \{(x, y) \in [s_1, s_3] \times [q(y_1, x), \ell_{am}(x)] \cup [s_3, s_4] \times [0, \ell_{am}(x)] \cup [s_4, s_6] \times [0, q_2(x)]\}.$$

Then

$$\begin{aligned} P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_1(\sqrt{3}/8)) &= \int_{T_s \setminus T(y_1, \sqrt{3}/8)} \frac{A(N_Y^r(x_1, \epsilon))}{A(T_\epsilon)^2} dy dx = \\ &= \int_{s_1}^{s_3} \int_{q(y_1, x)}^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx + \int_{s_3}^{s_4} \int_0^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx \\ &+ \int_{s_4}^{s_6} \int_0^{q_2(x)} \frac{A(\mathcal{P}(Q_1, N_1, N_2, Q_6))}{A(T_\epsilon)^2} dy dx = -\frac{r^4 - 2r^2 - 63}{676r^2}. \end{aligned}$$

where  $A(\mathcal{P}(Q_1, N_1, N_2, Q_6)) = \frac{\sqrt{3}}{576} (12rx + 4\sqrt{3}ry + 3)(12rx + 4\sqrt{3}ry - 3)$  is same as before.

Next,  $x_1$  between  $q_2(x)$  and  $\ell_s(x)$  is equivalent to  $x_1 \in R_2(\sqrt{3}/8)$  where

$$R_2(\sqrt{3}/8) = \{(x, y) \in [s_4, s_6] \times [q_2(x), \ell_{am}(x)] \cup [s_6, 1/2] \times [0, \ell_s(x)]\}.$$

Then

$$\begin{aligned} P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_2(\sqrt{3}/8)) &= \int_{s_4}^{s_6} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))}{A(T_\epsilon)^2} dy dx \\ &+ \int_{s_6}^{1/2} \int_0^{\ell_s(x)} \frac{A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6))}{A(T_\epsilon)^2} dy dx = \frac{64r^4 - 768r^3 + 1824r^2 - 128r - 1467}{2028r^2}. \end{aligned}$$

where  $A(\mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)) = \left(-\frac{\sqrt{3}}{12}y^2 - \frac{1}{2}xy - \frac{\sqrt{3}}{4}x^2\right)r^2 + \left(\frac{3}{4}y + \frac{3\sqrt{3}}{4}x\right)r - \frac{19}{64}\sqrt{3}$ .

Furthermore,  $x_1$  above  $q_3(x)$  is  $x_1 \in R_3(\sqrt{3}/8)$  where  $R_3(\sqrt{3}/8) = \{(x, y) \in [s_{10}, 1/2] \times [\ell_s(x), \ell_{am}(x)]\}$ .

Then

$$P(X_2 \in N_Y^r(X_1, \epsilon), X_1 \in R_3(\sqrt{3}/8)) = \int_{s_{10}}^{1/2} \int_{q_3(x)}^{\ell_{am}(x)} \frac{1}{A(T_\epsilon)} dy dx = \frac{8(4r^2 - 12r + 9)}{39r^2}.$$

Hence for  $r \in [3/2, 2)$ ,

$$\mu_S(r, \sqrt{3}/8) = 6 \left( \frac{61r^4 - 768r^3 + 3494r^2 - 5120r + 2466}{2028r^2} \right) = \frac{61r^4 - 768r^3 + 3494r^2 - 5120r + 2466}{338r^2}.$$

For  $r \geq 4$ , it is trivial to see that  $\mu(r) = 1$ .

To find the covariance, we need to find the possible types of  $\Gamma_1(x_1, N_Y^r, \epsilon)$  and  $N_Y^r(x_1, \epsilon)$  for  $r \in [1, 4)$ . The intersection points of  $\xi_j(r, x)$  with  $\partial(T(\mathcal{Y}))$  and  $\partial(R(y_j))$  for  $j = 1, 2, 3$ , i.e.  $G_1 - G_6$  and  $L_1 - L_6$  are same as before. Recall also  $M_C, M_1, M_2, M_3$  and  $y_1, y_2, y_3$ . Then  $\Gamma_1(x_1, N_Y^r, \epsilon)$  is a polygon whose vertices are a subset of the above points.

There are six cases regarding  $\Gamma_1(x_1, N_Y^r, \epsilon)$  and one case for  $N_Y^r(x_1, \epsilon)$ . Each case  $j$ , corresponds to the region  $R_j(\sqrt{3}/8)$  in Figure 45 where  $q(y_1, x), q_2(x), q_3(x), s_j$  for  $j = 1, 3, 4, 6, 10$  are same as before and  $q_4(x) = -\sqrt{3}(4x - r)/4$ ,  $q_5(x) = \sqrt{3}(2 - r)/4$  and  $s_2 = (r - 1)/4$ ,  $s_5 = (r^2 - 2r + 3)/(4r)$ ,  $s_7 = 3r/16$ ,  $s_8 = (r - 1)/2$ . (see Figure 45).

Then, for

$$\begin{aligned} x_1 &= (x, y) \in R_1(\sqrt{3}/8), \Gamma_1(x_1, N_Y^r) = \mathcal{P}(Q_1, G_2, G_3, G_4, G_5, Q_6) \\ x_1 &\in R_2(\sqrt{3}/8), \Gamma_1(x_1, N_Y^r) = \mathcal{P}(Q_1, G_2, G_3, M_2, L_4, L_5, M_3, Q_6) \end{aligned}$$



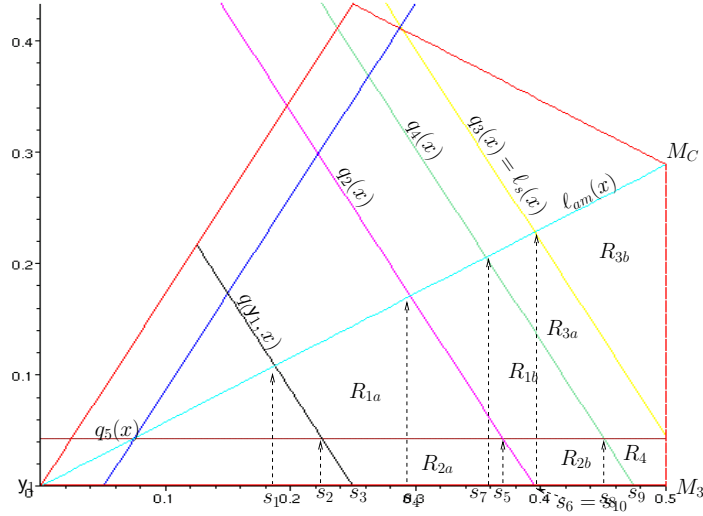


Figure 45: The regions corresponding to the seven cases for  $r \in [1, 4]$  with  $r = 1.9$

$$x_1 \in R_3(\sqrt{3}/8), \Gamma_1(x_1, N_y^r) = \mathcal{P}(G_1, G_2, G_3, G_4, G_5, G_6)$$

$$x_1 \in R_4(\sqrt{3}/8), \Gamma_1(x_1, N_y^r) = \mathcal{P}(G_1, G_2, G_3, G_4, G_5, G_6)$$

The explicit forms of  $R_j(\sqrt{3}/8)$ ,  $j = 1, \dots, 4$  are as follows:

$$R_1(\sqrt{3}/8) = \{(x, y) \in [s_1, s_2] \times [q(y_1, x), \ell_{am}(x)] \cup [s_2, s_7] \times [q_5(x), \ell_{am}(x)] \cup [s_7, s_8] \times [q_5(x), q_4(x)]\}$$

$$R_2(\sqrt{3}/8) = \{(x, y) \in [s_2, s_3] \times [q(y_1, x), q_5(x)] \cup [s_3, s_8] \times [0, q_5(x)] \cup [s_8, s_9] \times [0, q_4(x)]\}$$

$$R_3(\sqrt{3}/8) = \{(x, y) \in [s_7, s_8] \times [q_3(x), \ell_{am}(x)] \cup [s_8, 1/2] \times [q_5(x), \ell_{am}(x)]\}$$

$$R_4(\sqrt{3}/8) = \{(x, y) \in [s_8, s_9] \times [q_4(x), q_5(x)] \cup [s_9, 1/2] \times [0, q_5(x)]\}.$$

Let  $P_{2N}^r(\epsilon) := P(\{X_2, X_3\} \subset N_y^r(X_1, \epsilon))$ ,  $P_{2G}^r(\epsilon) := P(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon))$ , and  $P_M^r(\epsilon) := P(X_2 \in N_y^r(X_1, \epsilon), X_3 \in \Gamma_1^r(X_1, \epsilon))$ .

Now, by symmetry,  $P_{2N}^r(\sqrt{3}/8) = 6 P(\{X_2, X_3\} \subset N_y^r(X_1, \epsilon), X_1 \in T(y_1, \sqrt{3}/8))$ .

For  $r \in [7/4, 2)$ ,

$$P(\{X_2, X_3\} \subset N_y^r(X_1, \epsilon), X_1 \in T(y_1, \sqrt{3}/8)) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_y^r(x_1, \epsilon))^2}{A(T_\epsilon)^3} dy dx$$

$$= -\frac{261r^6 - 4608r^5 + 29105r^4 - 72960r^3 + 32575r^2 + 78848r - 67620}{65910r^2}.$$

where  $N_y^r(x_1, \epsilon) = \mathcal{P}(Q_1, N_1, N_2, Q_6)$  for  $x_1 \in R_{1a}(\sqrt{3}/8) \cup R_{2a}(\sqrt{3}/8)$ ,  $N_y^r(x_1, \epsilon) = \mathcal{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)$  for  $x_1 \in R_{1b}(\sqrt{3}/8) \cup R_{2b}(\sqrt{3}/8) \cup R_{3a}(\sqrt{3}/8) \cup R_4$ , and  $N_y^r(x_1, \epsilon) = T_\epsilon$  for  $x_1 \in R_{3b}(\sqrt{3}/8)$ , all of whose areas are given above. Hence for  $r \in [7/4, 2)$ ,

$$P(\{X_2, X_3\} \subset N_y^r(X_1, \epsilon)) = -\frac{261r^6 - 4608r^5 + 29105r^4 - 72960r^3 + 32575r^2 + 78848r - 67620}{10985r^2}.$$

Next, by symmetry,  $P_{2G}^r(\sqrt{3}/8) = 6 P(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon), X_1 \in T_s \setminus T(y_1, \sqrt{3}/8))$ , and

$$P(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon), X_1 \in T_s \setminus T(y_1, \sqrt{3}/8)) = \sum_{j=1}^4 P(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon), X_1 \in R_j(\sqrt{3}/8)).$$

For  $x_1 \in R_1 (\sqrt{3}/8)$ ,

$$\begin{aligned} P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon), X_1 \in R_1(\sqrt{3}/8)\right) &= \int_{s_1}^{s_2} \int_{q(y_1, x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx \\ &+ \int_{s_2}^{s_7} \int_{q_5(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx + \int_{s_7}^{s_8} \int_{q_5(x)}^{q_4(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx \\ &= \frac{18894 r^6 - 12248 r^5 - 131375 r^4 + 45360 r^3 + 584030 r^2 - 841816 r + 337155}{65910 r^4}. \end{aligned}$$

$$\text{where } A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon)) = \frac{\sqrt{3}(45r^2 + 32y\sqrt{3}x + 96x - 48x^2 - 80y^2 + 32\sqrt{3}y - 96)}{192r^2}.$$

For  $x_1 \in R_2 (\sqrt{3}/8)$ ,

$$\begin{aligned} P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon), X_1 \in R_2(\sqrt{3}/8)\right) &= \int_{s_2}^{s_3} \int_{q(y_1, x)}^{q_5(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx \\ &+ \int_{s_3}^{s_8} \int_0^{q_5(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx + \int_{s_8}^{s_9} \int_0^{q_4(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx \\ &= -\frac{4(4865 r^6 - 24063 r^5 + 49460 r^4 - 73210 r^3 + 98045 r^2 - 84107 r + 29010)}{32955 r^4}. \end{aligned}$$

$$\text{where } A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon)) = \frac{\sqrt{3}(96 + 32y\sqrt{3}x + 128\sqrt{3}ry - 224\sqrt{3}y - 192r + 96x + 93r^2 - 48x^2 + 176y^2)}{192r^2}.$$

For  $x_1 \in R_3 (\sqrt{3}/8)$ ,

$$\begin{aligned} P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon), X_1 \in R_3(\sqrt{3}/8)\right) &= \int_{s_7}^{s_8} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx + \int_{s_8}^{1/2} \int_{q_5(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx \\ &= -\frac{46293 r^6 - 100944 r^5 - 254880 r^4 + 506880 r^3 + 829440 r^2 - 2064384 r + 1048576}{98865 r^4}. \end{aligned}$$

$$\text{where } A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon)) = \frac{\sqrt{3}(3r^2 + 6x - 6x^2 - 6y^2 + 2\sqrt{3}y - 6)}{12r^2}.$$

For  $x_1 \in R_4 (\sqrt{3}/8)$ ,

$$\begin{aligned} P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, \epsilon), X_1 \in R_4(\sqrt{3}/8)\right) &= \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx + \int_{s_4}^{s_5} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon))^2}{A(T_\epsilon)^3} dy dx \\ &= \frac{8(3577 r^6 - 20548 r^5 + 45620 r^4 - 61760 r^3 + 79040 r^2 - 77824 r + 32256)}{32955 r^4}. \end{aligned}$$

$$\text{where } A(\Gamma_1(x_1, N_{\mathcal{Y}}^r, \epsilon)) = \frac{\sqrt{3}(-6r + 3x - 3x^2 + 5y^2 + 3r^2 + 3 + 4\sqrt{3}ry - 7\sqrt{3}y)}{6r^2}.$$

So

$$\begin{aligned} P_{2G}^r(\sqrt{3}/8) &= 6 \left( \frac{19032 r^6 - 243648 r^5 + 1118355 r^4 - 2085120 r^3 + 1534050 r^2 - 113664 r - 233639}{197730 r^4} \right) \\ &= \frac{19032 r^6 - 243648 r^5 + 1118355 r^4 - 2085120 r^3 + 1534050 r^2 - 113664 r - 233639}{32955 r^4}. \end{aligned}$$

Furthermore, by symmetry,

$$P_M^r(\sqrt{3}/8) = 6P\left(X_2 \in N_{\mathcal{Y}}^r(X_1, \epsilon), X_3 \in \Gamma_1^r(X_1, \epsilon), X_1 \in T_s \setminus T(y_1, \sqrt{3}/8)\right),$$

$$\begin{aligned} &P\left(X_2 \in N_{\mathbf{y}}^r(X_1, \epsilon), X_3 \in \Gamma_1^r(X_1, \epsilon), X_1 \in T_s \setminus T(y_1, \sqrt{3}/8)\right) \\ &= \sum_j P\left(X_2 \in N_{\mathbf{y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathbf{y}}^r), X_1 \in R_j\left(\sqrt{3}/8\right)\right) \end{aligned}$$
$$\begin{aligned}
R_{1a} \left( \sqrt{3}/8 \right) &= \{ (x, y) \in [s_1, s_2] \times [q(y_1, x), \ell_{am}(x)] \cup [s_2, s_4] \times [q_5(x), \ell_{am}(x)] \cup [s_4, s_5] \times [q_5(x), q_2(x)] \}, \\
R_{1b} \left( \sqrt{3}/8 \right) &= \{ (x, y) \in [s_4, s_7] \times [q_2(x), \ell_{am}(x)] \cup [s_7, s_5] \times [q_2(x), q_4(x)] \cup [s_5, s_8] \times [q_5(x), q_4(x)] \}, \\
R_{2a} \left( \sqrt{3}/8 \right) &= \{ (x, y) \in [s_2, s_3] \times [q(y_1, x), q_5(x)] \cup [s_3, s_5] \times [0, q_5(x)] \cup [s_5, s_6] \times [0, q_2(x)] \}, \\
R_{2b} \left( \sqrt{3}/8 \right) &= \{ (x, y) \in [s_5, s_6] \times [q_2(x), q_5(x)] \cup [s_6, s_8] \times [0, q_5(x)] \cup [s_8, s_9] \times [0, q_4(x)] \}, \\
R_{3a} \left( \sqrt{3}/8 \right) &= \{ (x, y) \in [s_7, s_{10}] \times [q_4(x), \ell_{am}(x)] \cup [s_{10}, s_8] \times [q_4(x), q_3(x)] \cup [s_8, 1/2] \times [q_5(x), q_4(x)] \}, \\
R_{3b} \left( \sqrt{3}/8 \right) &= \{ (x, y) \in [s_{10}, 1/2] \times [q_3(x), \ell_{am}(x)] \}.
\end{aligned}$$

For  $x_1 \in R_{1a}(\sqrt{3}/8)$ ,

For  $x_1 \in R_{1b} (\sqrt{3}/8)$ ,

For  $x_1 \in R_{2a}(\sqrt{3}/8)$ ,

For  $x_1 \in R_{2b}(\sqrt{3}/8)$ ,

$$\begin{aligned}
P\left(X_2 \in N_Y^r(X_1, \epsilon), X_3 \in \Gamma_1^r(X_1, \epsilon), X_1 \in R_{2b}\left(\sqrt{3}/8\right)\right) &= \int_{s_5}^{s_6} \int_{q_2(x)}^{q_5(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx \\
&+ \int_{s_6}^{s_8} \int_0^{q_5(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx + \int_{s_8}^{s_9} \int_0^{q_4(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx \\
&= \frac{2(r-2)(r^2-3)(467r^8-50r^7-7789r^6+1290r^5+16083r^4-4110r^3-7311r^2-810r+702)}{98865r^5}.
\end{aligned}$$

For  $x_1 \in R_{3a}(\sqrt{3}/8)$ ,

$$\begin{aligned} P\left(X_2 \in N_Y^r(X_1, \epsilon), X_3 \in \Gamma_1^r(X_1, \epsilon), X_1 \in R_{3a}(\sqrt{3}/8)\right) &= \int_{s_7}^{s_{10}} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx \\ &+ \int_{s_{10}}^{s_8} \int_{q_4(x)}^{q_3(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx + \int_{s_8}^{1/2} \int_{q_5(x)}^{q_3(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx \\ &= \left[3934 r^{12} - 11040 r^{11} - 2352 r^{10} - 283680 r^9 + 1239855 r^8 - 751008 r^7 - 3225344 r^6 \right. \\ &\quad \left. + 6125568 r^5 - 3847680 r^4 + 81920 r^3 + 1843200 r^2 - 2045952 r + 905472\right] / \left[395460 r^6\right]. \end{aligned}$$

For  $x_1 \in R_{3b}(\sqrt{3}/8)$ ,

$$\begin{aligned} P\left(X_2 \in N_Y^r(X_1, \epsilon), X_3 \in \Gamma_1^r(X_1, \epsilon), X_1 \in R_{3b}(\sqrt{3}/8)\right) &= \\ \int_{s_{10}}^{1/2} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx &= \frac{64(2r^4 - 4r^2 + 4r - 3)(2r - 3)^2}{507 r^6}. \end{aligned}$$

For  $x_1 \in R_4(\sqrt{3}/8)$ ,

$$\begin{aligned} P\left(X_2 \in N_Y^r(X_1, \epsilon), X_3 \in \Gamma_1^r(X_1, \epsilon), X_1 \in R_4(\sqrt{3}/8)\right) &= \\ \int_{s_8}^{s_9} \int_{q_4(x)}^{q_5(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx + \int_{s_9}^{1/2} \int_0^{q_5(x)} \frac{A(N_Y^r(x_1, \epsilon)) A(\Gamma_1(x_1, N_Y^r, \epsilon))}{A(T_\epsilon)^3} dy dx &= \\ = \frac{8(48r^6 - 22r^5 - 73r^4 - 3114r^3 + 4391r^2 + 1574r - 2850)(r - 2)^2}{32955 r^2}. \end{aligned}$$

Then

$$\begin{aligned} P_M^r(\epsilon) &= 6 \left( - \left[ 49r^{12} - 1536r^{11} + 17952r^{10} - 129280r^9 + 609420r^8 - 1728768r^7 + 2757670r^6 \right. \right. \\ &\quad \left. \left. - 3013632r^5 + 3418140r^4 - 3829760r^3 + 3026880r^2 - 1571616r + 445284 \right] / \left[ 395460r^6 \right] \right) \\ &= - \left[ 49r^{12} - 1536r^{11} + 17952r^{10} - 129280r^9 + 609420r^8 - 1728768r^7 + 2757670r^6 \right. \\ &\quad \left. - 3013632r^5 + 3418140r^4 - 3829760r^3 + 3026880r^2 - 1571616r + 445284 \right] / \left[ 65910r^6 \right]. \end{aligned}$$

So

$$\begin{aligned} \mathbf{E}_{\sqrt{3}/8}^S[h_{12} h_{13}] &= - \left[ 49r^{12} - 1536r^{11} + 18735r^{10} - 143104r^9 + 677703r^8 - 1704000r^7 + 1737040r^6 \right. \\ &\quad \left. - 691968r^5 + 1681230r^4 - 3716096r^3 + 3260519r^2 - 1571616r + 445284 \right] / \left[ 3295r^6 \right]. \end{aligned}$$

Hence

$$\begin{aligned} \nu_S\left(r, \sqrt{3}/8\right) &= \\ &- \left[ 637r^{12} - 19968r^{11} + 299370r^{10} - 3265792r^9 + 24051519r^8 - 112023360r^7 + 328179640r^6 \right. \\ &\quad \left. - 602490624r^5 + 673558110r^4 - 427086848r^3 + 133604087r^2 - 20431008r + 5788692 \right] / \left[ 428415r^6 \right]. \end{aligned}$$

Derivation of  $\mu_S(r, \epsilon)$  and  $\nu_S(r, \epsilon)$  for segregation with  $\epsilon = \sqrt{3}/4$  and with  $\epsilon = 2\sqrt{3}/7$  are similar.

## Appendix 5: Proof of Corollary 1

In the multiple triangle case,

$$\mu(r, J) = \mathbf{E}[\rho_n(r, J)] = \frac{1}{n(n-1)} \sum_{i < j} \mathbf{E}[h_{ij}] = \frac{1}{2} \mathbf{E}[h_{12}] = \mathbf{E}[\mathbf{I}(A_{12})] = P(X_2 \in N_Y^r(X_1)).$$

By definition of  $N_{\mathcal{Y}}^r(\cdot)$ ,  $P(X_2 \in N_{\mathcal{Y}}^r(X_1)) = 0$  if  $X_1$  and  $X_2$  are in different triangles. So by the law of total probability

$$\begin{aligned}\mu(r, J) &:= P(X_2 \in N_{\mathcal{Y}}^r(X_1)) = \sum_{j=1}^J P(X_2 \in N_{\mathcal{Y}}^r(X_1) \mid \{X_1, X_2\} \subset T_j) P(\{X_1, X_2\} \subset T_j) \\ &= \sum_{j=1}^J \mu(r) P(\{X_1, X_2\} \subset T_j) \quad (\text{since } P(X_2 \in N_{\mathcal{Y}}^r(X_1) \mid \{X_1, X_2\} \subset T_j) = \mu(r)) \\ &= \mu(r) \sum_{j=1}^J \left[ A(T_j) / A(C_H(\mathcal{Y})) \right]^2 \quad (\text{since } P(\{X_1, X_2\} \subset T_j) = (A(T_j) / A(C_H(\mathcal{Y})))^2)\end{aligned}$$

Then  $\mu(r, J) = \mu(r) \cdot \left( \sum_{j=1}^J w_j^2 \right)$  where  $\mu(r)$  is given in Equation (8).

Furthermore, the asymptotic variance is

$$\begin{aligned}\nu(r, J) &= \mathbf{E}[h_{12} h_{13}] - \mathbf{E}[h_{12}] \mathbf{E}[h_{13}] \\ &= P(\{X_2, X_3\} \subset N_{\mathcal{Y}}^r(X_1)) + 2P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r)) \\ &\quad + P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r)) - 4(\mu(r, J))^2.\end{aligned}$$

Let  $P_{2N}^r := P(\{X_2, X_3\} \subset N_{\mathcal{Y}}^r(X_1))$ ,  $P_{2G}^r := P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r))$ , and  $P_M^r := P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r))$ . Then for  $J > 1$ , we have

$$\begin{aligned}P(\{X_2, X_3\} \subset N_{\mathcal{Y}}^r(X_1)) &= \sum_{j=1}^J P(\{X_2, X_3\} \subset N_{\mathcal{Y}}^r(X_1) \mid \{X_1, X_2, X_3\} \subset T_j) P(\{X_1, X_2, X_3\} \subset T_j) \\ &= \sum_{j=1}^J P_{2N}^r (A(T_j) / A(C_H(\mathcal{Y})))^3 = P_{2N}^r \left( \sum_{j=1}^J w_j^3 \right).\end{aligned}$$

Similarly,  $P(X_2 \in N_{\mathcal{Y}}^r(X_1), X_3 \in \Gamma_1(X_1, N_{\mathcal{Y}}^r)) = P_M^r \left( \sum_{j=1}^J w_j^3 \right)$  and  $P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{\mathcal{Y}}^r)) = P_{2G}^r \left( \sum_{j=1}^J w_j^3 \right)$ , hence,

$$\nu(r, J) = (P_{2N}^r + 2P_M^r + P_{2G}^r) \left( \sum_{j=1}^J w_j^3 \right) - 4\mu(r, J)^2 = \nu(r) \left( \sum_{j=1}^J w_j^3 \right) + 4\mu(r)^2 \left( \sum_{j=1}^J w_j^3 - \left( \sum_{j=1}^J w_j^2 \right)^2 \right),$$

so conditional on  $\mathcal{W}$ , if  $\nu(r, J) > 0$  then  $\sqrt{n}(\rho_n(r, J) - \mu(r, J)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu(r, J))$ . ■